



UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Tres modelos contemporáneos de agencia humana. Un estudio sobre la motivación y la deliberación moral

Autora:

Leticia Elena Naranjo Gálvez

Director:

Dr. D. Carlos Thiebaut Luis-André

**Facultad de Humanidades, Comunicación y Documentación
Departamento de Humanidades: Filosofía, Lenguaje y Literatura**

Getafe, Noviembre de 2009

TESIS DOCTORAL

Tres modelos contemporáneos de agencia humana. Un estudio sobre la motivación y la deliberación moral

Autora: Leticia Elena Naranjo Gálvez

Director: Dr. D. Carlos Thiebaut Luis-André

Firma del Tribunal Calificador:

Firma

Presidente: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Secretario: (Nombre y apellidos)

Calificación:

Getafe, de de 2010

Agradecimientos

Para la elaboración del presente trabajo conté con el apoyo económico de una beca de investigación otorgada por el Departamento de Humanidades, Filosofía, Lenguaje y Literatura, de la Universidad Carlos III de Madrid, entre los años 2006 y 2009, y de un préstamo-beca de la Universidad del Rosario (de Colombia) entre 2003 y 2005.

Esta tesis no se hubiera podido terminar ni entregar sin la ayuda amorosa e incondicional de mi esposo, Alexander, a quien sencillamente no tengo cómo agradecer todo lo que ha hecho por mí.

En la etapa más crítica de la elaboración de este trabajo conté con la acogida y la ayuda de mis padres, María Victoria y Jairo, y de mis suegros Rosmira y Juan, así como con el apoyo fraterno e invaluable de mi amigo del alma, Wilson Herrera, de su esposa Camila De Gamboa y de su madre, Beatriz Tapias. Con Camila y Wilson, además, estoy en deuda por lo mucho que he aprendido de ellos, tanto, que parte de lo que pueda aportar esta tesis, se debe a su influencia. Mis hermanos Ximena y Rodrigo, mi tía Rosalía, y mis amigas Inés Calderón y Claudia Ruiz siempre me animaron desde lejos.

Durante mis años en España conté con el apoyo de tantas personas que espero que mi flaca memoria no traicione la enorme gratitud que siento por cada una de ellas. Agradezco a mi cuñada Diana y a su hija Alejandra, así como a nuestros amigos Herman Berrío y Claudia Torres, su afecto y su acogida.

También me siento en deuda con nuestros amigos y compañeros de fatigas, con quienes nos dábamos ánimos mutuamente, con los que compartimos penas y alegrías pero, sobre todo, unas conversaciones que, amén de alegrarnos, de enriquecernos y de abrirnos horizontes, nos sirvieron de acicate y de descanso a Alexander y a mí: Estela Meza y Jorge Rodríguez, Encarna López y Rosa Torralba, Yanitza Giraldo y Edwin Laniado, Esperanza Díaz y Pedro García, Mauricio Lombana, Guillermo De Eugenio y Diana Greño, María Santos Hernández y Juan Mechero, Regina Sonet y Daniel González, Daniel Martínez y Bogdan Marola, Gilmer Alarcón y Tatiana Rincón, Ana Gómez y Omar Molina, Johana Salge y Bernhard Schima, Elsa Fajardo y Thierry Lodomez. Nos hubiéramos sentido muy solos sin ellos.

A mis compañeros de despacho, Elena Azofra y Alfredo Kramarz, les agradezco su simpatía y su espíritu de servicio.

A la Doctora María Antonia Hernández y a Martha Camargo, les agradezco las luces y el consuelo. A Dayana Nieto y Enrique Hernández, y a las hermanas Barberena-Nisimblat, van mis agradecimientos por su solidaridad y su acogida en los momentos más oportunos.

De una manera especial quisiera agradecer al Profesor Carlos Thiebaut lo mucho que me ha enseñado, el tiempo y el cuidado que dedicó a leer lo que escribía, así como su paciencia y el interés que puso en todo este proceso. Los aciertos que pueda tener de esta tesis -y ninguno de sus fallos- se deben a él y a lo mucho que me ayudó.

También agradezco a los Profesores Fernando Broncano y Antonio Gómez lo que aprendí de ellos, así como su actitud siempre afable y generosa.

Todas y todos han sido imprescindibles, y espero no haber olvidado a ninguno de los que me aportaron y me hicieron más fácil el camino.

-Norah: Se han cometido muchos errores conmigo. En primer lugar, por parte de papá y, luego, por parte tuya. Cuando vivía con papá, él me manifestaba todas sus ideas, y yo las seguía. Si tenía otras diferentes, me guardaba muy bien de decirlo, porque no le habría gustado. Me llamaba su muñequita, y jugaba conmigo, como yo con mis muñecas. Después vine a esta casa contigo . . . Pasé de manos de papá a las tuyas. Tú me formaste a tu gusto y yo participaba de él . . . o lo fingía . . . creo que lo uno y lo otro . . . Vivía de hacer piruetas para divertirme, como tú querías. Tú y papá habéis cometido un grave error conmigo: sois culpables de que no haya llegado a ser nunca nada.

-Torvald: Pero qué injusta y desagradecida eres, ¿no has sido feliz aquí?

-Norah: No, nunca. Creí serlo, pero no lo he sido jamás . . . sólo estaba alegre y eso es todo. Eras tan bueno conmigo . . . Pero nuestro hogar no ha sido más que un cuarto de recreo. He sido muñeca grande en esta casa, como fui muñeca pequeña en casa de papá. Y a su vez los niños han sido mis muñecos.

-Torvald: Hay algo de verdad en lo que dices . . . aunque muy exagerado. Pero desde hoy todo cambiará; ya han pasado los tiempos de jugar y ha llegado la hora de la educación.

-Norah: La educación de quién, ¿la mía o la de los niños?

-Torvald: La tuya y la de los niños.

-Norah: Ay, Torvald, tú no eres capaz de educarme, y yo ¿qué preparación tengo para educar a los niños? Es una labor superior a mis fuerzas. Hay otra de la que debo ocuparme antes. Debo procurar educarme a mí misma. Tú no eres capaz de ayudarme en esta tarea y por ello necesito estar sola. Y por esa razón voy a dejarte.

-Torvald: ¿Qué dices!

-Norah: Necesito estar sola para orientarme sobre mí misma y sobre lo que me rodea. No puedo quedarme más contigo.

-Torvald: ¡Qué horror! Traicionar así los deberes más sagrados . . .

-Norah: Tengo otros deberes no menos sagrados.

-Torvald: ¿Qué deberes son esos?

-Norah: Mis deberes conmigo misma.

-Torvald: Ante todo eres esposa y madre.

-Norah: Ante todo soy un ser humano, igual que tú . . . o al menos debo intentar serlo. Sé que la mayoría de los hombres y los libros te darán la razón. Pero ahora no puedo conformarme con lo que dicen hombres y libros. Tengo que pensar por mi cuenta en todo esto y tratar de comprenderlo.

-Torvald: ¿Es que no te basta con tu puesto en este hogar? ¿No tienes una guía infalible para estos dilemas, no tienes la religión?

-Norah: Ay, Torvald, no sé qué es la religión. Sólo sé lo que me dijo el pastor Hensen cuando me preparaba para la confirmación . . . cuando esté sola y libre examinaré también ese asunto . . .

-Torvald: Hablas como una niña. No comprendes nada de la sociedad en que vivimos.

Norah: No, seguro que no. Pero ahora quiero tratar de comprenderlo y averiguar a quién asiste la razón, si a la sociedad o a mí.

(H. Hibsén)

Índice

Introducción	7
Primera Parte: La imbricación entre agencia moral “reducida” y racionalidad maximizadora: la propuesta de Gauthier	17
Capítulo 1: El nexo entre moral y autointerés: los supuestos iniciales de <i>La moral por acuerdo</i>	21
1. El contrato, la imparcialidad y el rechazo de la moral “de fines”. El tipo de agente que surge de este primer esquema-----	22
1.1 El uso de la noción de imparcialidad -----	25
1.2 El agente “no-concernido” y sus razones pre-morales para actuar moralmente. El mercado como modelo de moralidad -----	28
1.3 El maximizador irrestricto MI vs. el cooperador o maximizador constreñido MR. El problema de la motivación moral de este último -----	31
1.4 El modelo mecanicista, la moral como anomalía y la moderna sociedad de mercado como “caso ejemplar”-----	34
2. El modelo de la teoría de la elección racional y la reforma normativa que propone Gauthier-----	37
2.1 El esquema de la teoría de la decisión-----	37
2.2 Las preferencias cualificadas o “consideradas” y el problema de lo normativo -----	43
2.3 La noción de valor y el subjetivismo de los valores defendido por Gauthier. La tensión entre dicho subjetivismo y el interés por lo normativo -----	47
3. La “arena de la moral”: cooperación, autointerés y agencia económica-----	53
3.1 El mercado, sus agentes egoístas y por qué la moral reemplaza la mano invisible. Las tensiones entre el mercado sin fallos y los mercados reales-----	55
3.2 El modelo de Robinson Crusoe. El agente económico: los problemas para su autonomía y moralidad -----	61
3.3 El carácter ejemplar de la sociedad de mercado. La fusión entre agencia moral y agencia económica -----	68

Capítulo 2: Cómo unir egoísmo y cooperación: una teoría de la justicia que muestre a la moral como la solución a los fallos del mercado 81

1. La necesidad del contrato social ante la escasez y el autointerés como “circunstancias de la justicia”-----82
 - 1.1 Algunas dudas sobre la insistencia inicial en el modelo del mercado y un posible desliz de lo político a lo económico-----85
 - 1.2 El principio del *minimax relative concession* MRC y la negociación en busca de un acuerdo *justo* y racional. La elección de una estrategia conjunta (*joint*) y, de nuevo, el problema de la motivación *moral* -----88
 - 1.3 Las razones para la justicia o para cumplir el pacto social; las respuestas de Hobbes y de Gauthier al “insensato”. Cómo sobrevivir en una sociedad de MMII si se es un MR, y el problema del (no)carácter moral de los supervivientes -----91
2. Un primer paso para la moralización del *homo oeconomicus*: su elección paramétrica de una disposición a actuar y su participación en el acuerdo para el contrato social -----98
 - 2.1 El difícil equilibrio entre tres críticas: contra al emotivismo de Parfit, contra los “neokantianos” y contra la idea de una racionalidad no-maximizadora ----- 100
 - 2.2 El *proviso* lockeano, la “igualdad en la racionalidad”, y cómo adaptar –en clave maximizadora– el rawlsiano punto de vista “arquimédico” para el contrato social ----- 103
3. Las exigencias normativas del *proviso*, y las tensiones entre lo aplicable y lo ideal del modelo de Gauthier----- 114
 - 3.1 El aspecto “aplicable-realista” de su modelo, según Gauthier----- 115
 - 3.2 El mercado como el resultado, tanto ideal-conceptual, como histórico, de la aplicación del *proviso*----- 121

Capítulo 3: La conclusión del proceso de conversión del *homo oeconomicus* en el individuo liberal..... 129

1. Las sospechas del sofista: la justicia concebida como “mal necesario” ----- 130
 - 1.1 La idea de un hombre “natural” y pre-moral----- 133
 - 1.2 Para salvar al hombre natural: motivaciones “primarias” vs. “secundarias”. De nuevo, el *homo oeconomicus* como talón de Aquiles de la propuesta de Gauthier----- 137
2. El tercer paso en la moralización del *homo oeconomicus*: su transformación en el individuo liberal ----- 143
 - 2.1 La necesidad de “pacificar” el mercado----- 148

2.2	El aspecto institucional, la persistencia del egoísmo y las desigualdades sociales: ¿la autonomía en riesgo?-----	155
2.3	Racionalidad “meramente instrumental” vs. racionalidad “completa”; sociedad “artificialmente” justa vs. sociedad “esencialmente” justa -----	159
2.4	El valor “adaptativo” e instrumental de los sentimientos morales y del gusto por la socialización ¿Un posible sesgo reduccionista y elitista en Gauthier? -----	162
2.5	La prelación del fundamento maximizador frente a los sentimientos morales ¿Cómo asegurar una auténtica motivación moral para el “nuevo” <i>homo oeconomicus</i> ? -----	166
3.	La insalvable condición de escasez y sus ineludibles consecuencias: maximización y moralidad -----	170
3.1	La imposibilidad de la utopía marxista. Nuestra esencia maximizadora y la necesidad de la moral no sólo como restricción, sino también como medio -----	173
3.2	El irrenunciable individualismo, el valor instrumental de la cooperación y, de nuevo, la base racional de los sentimientos morales -----	176
3.3	La crítica a los yoes discontinuos de Parfit. La capacidad crítica del agente liberal y su preocupación por sus yoes futuros-----	182
3.4	El retrato definitivo del liberal ¿“Libre afectividad” vs. cohesión social? ¿Autonomía vs. condicionamientos sociales? -----	187
4.	Conclusiones parciales -----	195

Segunda Parte: Hacia un modelo más complejo de agencia: de la “persona” de Frankfurt al “evaluador fuerte” de Taylor..... 202

Capítulo 4: Frankfurt: la persona, su estructura volitiva y sus “preocupaciones” 206

1.	Complejidad volitiva y deseos “de segundo orden”-----	208
1.1	Deseos “efectivos”, voliciones de segundo orden y el caso de los <i>wantons</i> -----	212
1.2	Preocupaciones, identificación, “unidad” de agencia y libertad de la voluntad-----	215
2.	La “necesidad volitiva” y lo “impensable”-----	220
2.1	El <i>caring</i> : libertad y necesidad -----	221
2.2	La racionalidad prediscursiva, el amor y la exclusión de lo normativo-----	225
3.	La “jerarquía” de los deseos y el problema de la “identificación”-----	230
3.1	¿Una cadena infinita de deseos y una decisión inexplicable? -----	231

3.2	Una propuesta para fortalecer el agente “integrado” de Frankfurt, en contraste con el maximizador de Gauthier -----	239
Capítulo 5: El “evaluador fuerte” de Taylor		248
1.	Con Frankfurt pero más allá de él: la importancia del asunto valorativo-----	254
1.1	Evaluaciones “débiles” vs. evaluaciones “fuertes” -----	255
1.2	La idea de “forma de vida” -----	260
1.3	El lenguaje de las evaluaciones fuertes -----	265
1.4	El lenguaje del evaluador fuerte vs. el del maximizador. Lo cualitativo y lo cuantitativo-----	266
2.	Las formas de autointerpretación y la pregunta por el “yo” (<i>self</i>) que creemos ser-----	274
2.1	Ser un evaluador fuerte o ser un simple “sopesador” (<i>weigher</i>) de alternativas ----	275
2.2	La persona y el <i>wanton</i> de Frankfurt; el evaluador fuerte y el evaluador débil de Taylor-----	281
3.	Los dilemas morales y la imposibilidad del “yo desvinculado”-----	283
3.1	El problema de la elección “radical” -----	285
3.2	La libertad del agente <i>à la</i> Gauthier vs. las ataduras del evaluador fuerte. De la elección radical de Sartre a lo “impensable” de Frankfurt-----	292
4.	A modo de conclusión: los límites de la autoaclaración y la esperanza en lo dialógico--	298
Conclusiones.....		308
Bibliografía.....		324

Introducción

El presente trabajo tuvo como motivación original una preocupación por la influencia que a nivel del imaginario popular, pero también en el ámbito de las discusiones públicas y, en ocasiones, en los espacios académicos, parece tener un estereotipo, una caricatura que, aun a pesar de su carácter ficticio, ha terminado por ser tomada bastante en serio en algunos sectores de nuestra sociedad: el *homo oeconomicus*. Creo que este estereotipo resulta perturbador, sobre todo si se tiene en cuenta su influencia en una sociedad que, como la nuestra, debe buena parte de su autoimagen a dos ideales emancipatorios, estrechamente relacionados entre sí, y que parecen ser contradichos por el *homo oeconomicus*: el ideal de un sujeto autónomo y el de una sociedad libre de toda forma de sujeción. El primero de ellos es contradicho por este estereotipo, pues un sujeto pensado únicamente como agente económico estaría lejos de ser el individuo que es capaz de reflexionar sobre sí mismo y sobre su entorno social. Por el contrario, un agente¹ autónomo sería reflexivo incluso hasta el punto de ponerse por encima de sus propios deseos y, gracias a esas capacidades reflexivas, podría tomar una distancia crítica que le permitiría juzgar y decidir, sin presiones externas y sin compulsiones internas, sobre el curso que quisiera darle a su vida, sobre aquellos fines o ideales a los que quisiera apuntar con ella, y sobre el tipo sociedad que quisiera construir con otros agentes. Lo cual significaría, entonces, que sus vínculos con ellos estarían signados por la libertad, por el mutuo reconocimiento, y no por la fuerza, ni la manipulación, ni la mutua instrumentalización.

¹ Aquí se entenderá al “agente” y al “paciente” bajo un esquema aristotélico de categorías, según el cual al primero se lo considera como un sujeto capaz de actuar y decidir por sí mismo, o como moviéndose a la acción por sí mismo, de tal manera que dicha acción tiene su principio en el propio agente. Por el contrario, el “paciente” sería pensado como un ser que es objeto de la acción de otros, o de fuerzas que escapan a su poder, siendo empujado por dichas fuerzas o voluntades a las que no puede resistirse y de las que no es responsable, puesto que el principio de lo que “parece” que hace el paciente, realmente no está en él, sino en dichas fuerzas externas. Esta distinción aristotélica, su vez, se encuentra ligada a las oposiciones establecidas por la filosofía de la acción del siglo XX: “actuar” vs. “padecer”; un mero “cambio” o “movimiento” vs. una “acción” propiamente dicha. Por otra parte, estas categorías filosóficas son completamente ajenas a los términos económicos “principal” y “agente”, entendiéndose por el segundo aquel que opera con el fin de obtener beneficios para el primero →v.g. el empleado que produce para el empleador; el directivo que produce para el accionista—. Cf. Aristóteles. *Organon-Tratados de Lógica*. L. I: Categorías; Davidson, Donald (2001).

Esto último apunta al segundo ideal contradicho por el estereotipo del *homo oeconomicus*: el de la libertad de la que deberían gozar unos sujetos autónomos, cuyos vínculos no sean los de la fuerza, ni el engaño, ni ninguna otra forma de manipulación o coerción. Empero, estas características ideales no podrían ser atribuidas al tipo de relaciones en las que se piensa involucrado al *homo oeconomicus*, las cuales son básicamente relaciones de mutua instrumentalización, incluso si a dicha instrumentalización se la concibe como limitada por ciertas reglas de juego, v.g. las del mercado perfectamente competitivo, en el que teóricamente no se daría ninguna forma de explotación o coerción. Al contrario de lo que sucede con el sujeto autónomo, al *homo oeconomicus* se lo piensa como una suerte de máquina maximizadora que es capaz de instrumentalizarlo todo, incluyendo sus relaciones con otros agentes, en aras de la satisfacción de unas preferencias que el sujeto no evalúa ni cuestiona, sino que se limita a satisfacer de la manera más eficiente. No en vano a este personaje se lo asocia con el “aplanamiento” de los valores que parece surgir de la globalización de los mercados y de la intrusión de la lógica económica en las diversas esferas de la vida humana. Se trataría, pues, de un sujeto que es reflexivo en un sentido mínimo, puesto que no cuestiona los fines o propósitos que persigue con sus acciones, si bien despliega una forma de “astucia” y, en este sentido, es pensado como dotado de ciertas herramientas racionales y deliberativas que le permiten encontrar los medios para el logro de tales propósitos.

El *homo oeconomicus* es concebido como una suerte de agente abstracto, sin historia, sin vínculos con otros agentes, sin cultura y sin compromisos. Lo cual también conlleva la idea de unos extraños vínculos sociales, ya que al *homo oeconomicus* se lo piensa como un ser que se relaciona con otros únicamente en tanto que propiciadores u obstáculos de sus planes. De modo que sus lazos con esos otros pueden ser bastante inestables, siendo controlados únicamente por los poderes disuasorios de las fuerzas del mercado o de las leyes que pauten las relaciones comerciales o contractuales, si bien, por supuesto, dichas leyes son sólo eso: restricciones con las que se cuenta estratégicamente, pero no un vinculante objeto de compromiso. ¿Qué podría ser lo aterrador de esta caricatura, además del hecho preocupante de que, a pesar de ser una ficción, sin embargo parece que ha calado tan hondo en nuestra autoimagen como occidentales modernos? Pienso que parte de la respuesta está en que en ella se ilustra

una idea empobrecida o reducida de la agencia humana, una idea que implica la imposición de restricciones a las capacidades que quisiéramos atribuir a un agente humano reflexivo y autónomo, restricciones que significarían una merma en su autoconciencia y en su libertad, así como un deterioro de las relaciones que puedan establecer entre sí los agentes humanos pensados bajo este modelo del *homo oeconomicus*. Pues este maximizador egoísta no se cuestiona a sí mismo, no reflexiona sobre sus propósitos, no evalúa sus relaciones con otros agentes y, para sus deliberaciones, sólo toma en cuenta un repertorio bastante limitado de elementos de juicio, dado que, asimismo parece que el campo de su experiencia moral sería bastante restringido.

Desde la “lógica” que impone este estereotipo, todo asunto o problema que sea inanalizable desde este estrecho repertorio de recursos para la deliberación, es considerado ininteligible o inabordable por la “razón”. Así, podemos llegar al curioso resultado de que aquello que es visto como máximamente “racional” por parte de quienes se tomen en serio la figura del *homo oeconomicus*, termina por tener un revés, también aterrador. A saber, una suerte de reacción o de réplica en contra de lo que dicha figura representa: los ímpetus de aquellos otros que, por el contrario, vean en la actitud del *fanático* –v.g. el fanático nacionalista, o el fanático religioso– la expresión de una “liberación” de la falta de valores y de la impersonalidad de la máquina económica. Parecería que el intento de racionalizar la vida mediante una forma excluyente de racionalidad, termina por constituirse en una forma de irracionalismo que, a su vez, propiciaría la emergencia de otras formas de irracionalismo en las que también se darían formas excluyentes de racionalidad, v.g. la del fanático que no cuestiona la causa a la que sirve, y que está dispuesto a sacrificarlo todo por ella, incluyendo a otros agentes o a sí mismo.

La preocupación que pueda generar el ejercicio de formas reducidas y excluyentes de racionalidad, como aquella que se ejemplifica en la caricatura del *homo oeconomicus*, estaría animando el propósito del presente trabajo: llevar a cabo una indagación sobre las posibilidades de una agencia humana a la que se pueda pensar, por el contrario, como no restringida, como no limitada por formas excluyentes de deliberación. Así mismo, se trataría de una forma de agencia asociada a una base motivacional que tampoco se restringiría a un único y opaco motor de las acciones

humanas, v.g. el mero autointerés del *homo oeconomicus*. Por tanto, se trataría de una base motivacional que estaría abierta a la evaluación, al tiempo que ésta implicaría una forma de agencia que no estaría imposibilitada para dirigirse a diversos objetos de reflexión, tales como la propia vida, o los fines e ideales del sujeto, así como sus relaciones con otros agentes.

Para llevar a cabo esta indagación, he partido del análisis de una propuesta filosófica en la que, a mi parecer, se expone de manera sólida la figura del *homo oeconomicus*, asumida como ideal de la agencia humana: la propuesta de David Gauthier, desarrollada en su obra *La moral por acuerdo*. Luego de identificar los contornos que presenta el agente modélico de Gauthier –su maximizador autointeresado–, intentaré mostrar de qué manera, si se examina la obra de otros dos filósofos contemporáneos, Harry Frankfurt y Charles Taylor, se podrían apreciar mejor, tanto los vacíos que cabría atribuir al modelo de agencia que parece obtenerse de la propuesta de Gauthier, como las posibles soluciones o aportes que, frente a dicho modelo, podrían ofrecerse desde los discursos de Frankfurt y de Taylor. La elección de estos autores se debe, en primer término a que, a mi entender, de ellos puede hacerse una lectura según la cual se partiría de una idea muy restringida de la agencia moral –expuesta por Gauthier– hasta llegar progresivamente –vía Frankfurt– a otra a la que podemos considerar más rica e incluyente –presentada por Taylor–. En segundo lugar, he elegido a estos autores por la importancia de sus aportes y la solidez de sus discursos, en cuanto a que cada uno de ellos, desde mi punto de vista, admitiría ser calificado como paradigmático o representativo de una cierta idea acerca de la agencia humana.

La primera parte de la tesis se divide en tres capítulos en los que se expone y analiza la propuesta de Gauthier en su obra *La moral por acuerdo*, con el fin de extraer de ella el modelo de agencia moral que de ella surge o que en ella se privilegia. Se verá que dicho autor nos muestra una pintura bastante limitada del agente, de su ejercicio de la racionalidad práctica, de las relaciones que establece con otros sujetos, de sus experiencias morales, y de la base motivacional que explica sus acciones en el campo moral. Pero, por otro lado, también habría que reconocer que tal modelo del agente humano –*homo oeconomicus*– es el que parece haberse impuesto sobre sus competidores, tanto a nivel de los discursos –filosófico, económico– como a nivel del imaginario popular de Occidente. Este modelo, a su vez, es indesligable del contexto en

el que se desarrollan las relaciones de este tipo de agente con sus congéneres: el mercado, que en Gauthier aparece como el escenario que provee la base normativa para los nexos que puedan establecerse entre los agentes, a quienes se conciben como no interesados en la suerte de sus semejantes y como motivados fundamentalmente por la búsqueda de su propio beneficio. Lo cual, para el autor, está estrechamente relacionado con un ideal moral de libertad representado en la figura de Robinson Crusoe.

He escogido a David Gauthier porque me parece un intento bastante serio, esforzado e ingenioso por lograr fundamentar la moral a partir de un modelo de racionalidad que, si bien podría ser señalado como “reducido”, también es cierto que, al haber sido objeto de un largo y cuidado trabajo de elaboración, aparece como una forma incontrovertible de racionalidad práctica: aquella que se aprecia en la teoría de la elección racional. El proyecto de fundamentación de Gauthier, a su vez, es indesligable de su modelo de agencia: ese Robinson al que se piensa como un maximizador egoísta pero prudente, que es capaz no sólo de perseguir de manera eficiente su propio beneficio, sino que igualmente es capaz de restringirse en dicha búsqueda cuando vea que el cumplimiento de reglas morales, también seguidas por los demás agentes, sea necesario para el logro de su propia utilidad. Gauthier, pues, representa, como bien dice Rawls, el mejor intento por hacer que lo razonable se reduzca a lo racional,² desarrollando un esquema sólido al cual vale la pena enfrentarse si se está preocupado por las consecuencias –a mi entender, indeseables– que dicha reducción traería para las posibilidades de la agencia humana, y para el ejercicio de la razón práctica; las cuales se limitarían a la estrechez de la agencia económica y de una forma meramente maximizadora de racionalidad.

Luego de examinar el discurso de Gauthier, se desarrollará la segunda parte de la tesis, que a su vez se divide en dos capítulos, dedicados respectivamente a las propuestas de Harry Frankfurt y de Charles Taylor. Este orden, como intentaré mostrar, obedece a que, como arriba se menciona, se podría ver en el modelo de agente que se obtiene del primero una suerte de transición, en cuanto a su complejidad, entre el modelo más restringido que aparece en Gauthier, y el modelo más incluyente o complejo que surge del discurso de Taylor. Así, en el cuarto capítulo, se analiza la propuesta de Frankfurt, en la cual nos encontramos, de un lado, con elementos de mayor

² Rawls, John (1996; conferencia II, § 1, pp. 69-70).

complejidad que, en mi opinión, implican un interesante salto cualitativo frente al agente modélico de Gauthier aunque, de otra parte, se siguen conservando otros aspectos que implican una suerte de continuidad con respecto al autor de *La moral por acuerdo*. En cuanto el primer tipo de elementos, en Frankfurt hallamos la novedad de un agente que puede evaluar sus deseos y que puede motivarse a actuar por asuntos que demandan de él una atención desinteresada, asuntos sobre los cuales la persona se encuentra *preocupada* o a los que concede una importancia que va más allá de la satisfacción de sus deseos. Adicionalmente, este autor, a diferencia de Gauthier, asume aquello que en Strawson y en Habermas³ sería considerado el punto de vista que debería asumir el filósofo moral que, en vez de objetivar la praxis moral de modo que ésta sólo se pueda explicar “desde fuera” de ella —desde una suerte de “verdadera realidad” de la cual la moral sólo podría ser un epifenómeno— asume la actitud “participativa” e intenta, más bien, dar cuenta de la realidad moral desde ella misma, desde las intuiciones y experiencias de los agentes morales. Con lo cual, a pesar de que Frankfurt insista en no utilizar el apelativo “moral” para referirse al campo de las “preocupaciones” humanas (*caring*), creo que nos hallaríamos lejos de un cierto compromiso reduccionista que puede atribuirse a Gauthier.

Por otra parte, el segundo tipo de elementos, aquellos en los que podemos ver en Frankfurt una suerte de continuidad con respecto a Gauthier, podrían ser sintetizados en dos aspectos: 1) La persistencia de la figura de un agente solitario, que aparece pensado casi que *in abstracto*, sin contexto y al que, por tanto, tampoco se le puede concebir como dando y escuchando razones en un diálogo con otros agentes, y 2) La insistencia en un irracionalismo ético, en una postura anticognitivist, relacionada con la ausencia de referentes normativos que permitan explicar mejor la evaluación de los deseos, y que también conecten al agente con aquello que podría contar como buenas razones a los ojos de otros agentes. Si bien, a diferencia de Frankfurt, Gauthier no suscribe en absoluto una postura anticognitivist, pienso que terminaría por pagar el precio de cierto irracionalismo, causado por el hecho de que los alcances de la razón práctica, tal y como él los concibe, se limitarían al cálculo de costes-beneficios, quedando, por ende, casi toda la experiencia moral desterrada al campo de lo irracional.

³ Strawson, Peter (1995); Habermas, Jürgen (1994a).

Finalmente, en el quinto capítulo, se analizará la propuesta de Taylor, en el que puede verse la solución a algunas de las aporías a las que se llega en el discurso de Frankfurt, gracias a que en el esquema del primero pueden incluirse y radicalizarse algunos de los aportes de este último. Concretamente, el agente modélico de Frankfurt, la “persona”, podría ser convertida en el “evaluador fuerte” de Taylor, por la vía de superar su falta de recursos para dar cuenta de sus decisiones, para argumentar de cara a otros agentes y, por ende, para que dichas decisiones incluyan, pero también excedan los límites de aquello que en Frankfurt se denomina lo “volitivo” y que este autor insiste en mantener al margen de lo moral, lo ético y lo normativo.

En este punto debo introducir un pequeño paréntesis. Frankfurt no hace una distinción entre “lo ético” y “lo moral”, entendiendo por ambos el mismo tipo de asunto: aquello que atañe a las *normas* que rigen las *relaciones entre* los agentes o que se refieren al modo en que las decisiones de un agente puedan afectar a sus congéneres. Para efectos de este trabajo, entenderé por “lo moral” esta definición de Frankfurt, pero haré la diferencia que establece Habermas,⁴ y en virtud de la cual “lo ético” se refiere sobre todo a aquellos valores que definen el ideal de vida lograda de los agentes. Sin embargo, en muchos apartes de la tesis no haré esta diferencia, dado que, como veremos, especialmente en la parte dedicada a Taylor, habría ocasiones en las cuales a los agentes morales se les plantean problemas éticos que se solapan con sus preocupaciones morales, siendo difícil establecer una frontera entre ambos asuntos.

En Taylor encontramos un tipo de agencia humana que es capaz de los logros que hallamos en Frankfurt pero que, además de éstos, puede incluir en sus deliberaciones aquello que el segundo insiste en excluir de eso a lo que denomina como las “preocupaciones” de las personas, y en virtud de la cuales estas últimas llevan a cabo la evaluación de sus deseos y la estructuración de su identidad. En el autor de *Las fuentes del yo* vemos superados los problemas causados por la insistencia de Frankfurt de excluir lo normativo de las preocupaciones de las personas, a favor de una prelación de aquello que este segundo autor entiende como lo “volitivo”, campo que Frankfurt escinde por completo de lo cognitivo, pagando el precio de un irracionalismo que puede ser superado en Taylor. Este último muestra que las deliberaciones humanas se dan sobre la base de un horizonte de valor y de sentido que cada agente supone, si bien

⁴ Habermas, Jürgen (1990).

dicho horizonte, en su papel de contexto o de trasfondo, no resulta inmediatamente claro para el agente. Esto implica que la persona no puede simplemente elegir aquello que ha de considerar prioritario o valioso, sino que antes que hablar de elección habría más bien que considerar la deliberación, tanto moral como éticamente relevante, como un proceso de aclaración o de autoaclaración, en el cual el sujeto se hace cargo de aquello que considera más importante, desvelándolo y evaluándolo, sobre todo en ocasiones en las que se enfrente a un dilema moral.

Para Taylor, el agente debe, entonces, ser pensado como un ser atado a una red de significados que la persona supone en sus juicios y elecciones, pero que requieren por su parte ser sacados a la luz, mediante un proceso reflexivo de evaluación y autoevaluación, en el cual el agente debe intentar determinar qué es lo más importante o lo más valioso para él, mediante razones que debería poder exponer ante esos otros con los cuales comparte su horizonte de valor o de sentido. Estamos, pues, ante una figura, la del “evaluador fuerte”, que presenta una complejidad significativa con respecto a la “persona” de Frankfurt, pero sobre todo, con relación al maximizador egoísta presentado por Gauthier. Sin embargo, en el último capítulo, dedicado a Taylor, se intentará mostrar que ser un evaluador fuerte no es un hecho seguro para todos, sino un difícil logro, un *telos* deseable, pero que también puede ser puesto en peligro por la buena o mala fortuna moral de las personas, sobre todo si se tiene en cuenta el peso que pueda tener la relación con sus otros, de quienes el agente puede recibir o no un reconocimiento, o con respecto de los cuales puede contar o no con las condiciones necesarias para que alcance a asumir frente a ellos una actitud crítica y un mayor despliegue de su autonomía.

No obstante esta última salvedad, también intentaré que puedan verse en Taylor algunos elementos que nos permitirían, en palabras de Nozick, “ir más allá de Hume”,⁵ en cuanto a la posibilidad de una razón práctica que sea mucho más que “la esclava de las pasiones”, entre otras razones, porque es posible que esta expresión implique también una escisión bastante cuestionable entre la razón y “su otro”, escisión⁶ que tal vez esté a la base de una idea demasiado restringida de la agencia humana, y por la que

⁵ Nozick, Robert (1995, p. 91).

⁶ Sobre las sospechas que pueden caer sobre esta escisión, véase la crítica que hace I. M. Young a las dicotomías propias de lo que ella, tomando prestada la expresión de Adorno, denomina como una “lógica de la identidad”. Cf. Young, I. M. (1990, pp. 96-121).

acaso terminemos por pagar el precio de abandonar los aspectos más importantes de la vida individual y colectiva a la irracionalidad de las “razones” de la fuerza o de los avatares históricos.

Primera Parte

La imbricación entre agencia moral “reducida” y racionalidad maximizadora: la propuesta de Gauthier

En los siguientes tres capítulos analizaré la propuesta de David Gauthier en su obra *Morals by Agreement*, con el fin de señalar en ella los aspectos que conformarían lo que a mi entender admite ser visto como el revelador retrato de un tipo de *agente*¹ moral al que me permito llamar “reducido”. Este apelativo se debe, en primer lugar, a que creo que el agente así modelado por el autor canadiense se caracteriza por hacer un uso exclusivo de cierta forma de *racionalidad* práctica a la que, como intentaré mostrar, también podría considerarse como reducida: la racionalidad de la eficiencia de medios a fines o la racionalidad estrictamente técnica-instrumental. En segundo lugar, las razones que dicho agente tendría para actuar conformarían una base *motivacional* que así mismo pienso que puede ser descrita como reducida, en tanto se trata de razones que colapsan finalmente en el mero autointerés.

Esta suerte de doble reducción explica, a su vez, una tercera, en la que, a mi modo de ver, confluyen las otras dos: en el modelo presentado por Gauthier la disposición que tenga el agente para con *sus congéneres* y, por ende, para guiar su conducta por normas *morales*² que le vinculen con ellos, se origina igualmente en una motivación autointeresada. A ésta se remitirían en última instancia las razones que podría válidamente aducir, según nuestro autor, cualquier agente moral que pretenda justificar sus acciones como racionales. Por tanto son este tipo de razones las únicas que legítimamente darían cuenta del uso correcto que, para Gauthier, debe dársele a la expresión “racionalidad práctica”. El filósofo canadiense sostiene que dicha racionalidad sólo puede ser entendida como aquella que está orientada a la maximización de la utilidad esperada por parte de agentes racionales; y “racionales” en el estricto sentido de la moderna teoría de la elección racional (*Rational Choice*

¹ Remito al lector a un pie de página anterior, en el que explico en qué sentido se utiliza aquí el término “agente”, cuestión que se retomará reiteradamente a lo largo de la tesis.

² De modo similar a lo que sucede con la noción de “agente”, me permito recordar que en un pie de página anterior se ha explicitado lo que aquí se entenderá por “moral” y por “ética”.

Theory).³ Esto es, se trata de agentes que con la elección de sus acciones solamente buscan satisfacer sus preferencias u obtener el logro eficiente y eficaz de sus fines, de una manera tal que cada objetivo conseguido por parte de un agente sea pagado por éste al menor coste que le sea posible.

El intento de Gauthier se centrará en llevar a cabo lo que él entiende como la única fundamentación propiamente racional de la moral, por la vía del mostrar que son este tipo de razones ancladas en el autointerés y no en otra forma de motivación, v.g. el altruismo o el compromiso con bienes que el agente juzgue como más importantes que su propio beneficio, las que explican el hecho de que individuos definidos como autointeresados tengan motivos –racionalmente justificados– para ser morales. Es decir, para restringirse en algunas ocasiones en la búsqueda de la satisfacción de sus preferencias y, con ello, no perseguir en el corto plazo o de un modo miope sus intereses egoístas. De este modo los agentes estarían dispuestos a aceptar algunos constreñimientos a su comportamiento maximizador, con el fin de beneficiarse en el mediano y largo plazo de los frutos de la cooperación social. Lo cual supone que dicha cooperación y, por tanto, la aceptación de los mencionados constreñimientos le resulten tan convenientes a cada individuo, como lo serían para aquellos otros agentes con los cuales se relacione. En resumen, la *moralidad*, en el argot de Gauthier, consiste únicamente en este especial tipo de restricciones al propio egoísmo; la *racionalidad práctica* no es otra que la racionalidad técnica-instrumental y, finalmente, la *motivación para ser morales* se reduce a una búsqueda prudente,⁴ a largo plazo, de la maximización del propio beneficio.

Una de las consecuencias que más me interesa resaltar del modelo reducido de agencia que ejemplifica un discurso como el de Gauthier, es que termina por considerarse como imposible e ilegítimo cualquier intento de *evaluación propiamente racional de las preferencias* de los agentes, en especial de aquellas a las que algunos

³ Véase las definiciones de “teoría de la elección racional” y “teoría de la decisión” en: Resnik (1998); Abell (1991), Baron (1993), y Schmitz (2006), así como en los clásicos de la *Rational Choice Theory*: Von Neumann y Morgenstern (1953), y Luce y Raiffa (1958).

⁴ Aquí utilizo el término “prudente” un sentido distinto la *phronesis* aristotélica. Como bien lo muestra Carlos Thiebaut (2005, pp. 75-111) en su texto sobre la prudencia, esta expresión ha sufrido un cambio semántico en virtud del cual ha terminado por designar, sobre todo en el ámbito anglosajón, la virtud del cálculo de las consecuencias de las propias acciones; lo cual no coincide del todo con lo que se entiende como *phronesis* en el contexto de la *Ética a Nicómaco*. Cf. Aristóteles, EN, 1140a25 y ss.

autores designarían como preferencias “fuertes”;⁵ esto es, el tipo de preferencias en las que los individuos manifiestan sus compromisos éticos o valorativos más fundamentales. Con este resultado y a la vista de las fisuras encontradas en el mencionado modelo, espero que a partir del cuarto y, sobre todo del quinto capítulo, se pueda apreciar con mayor claridad que resulta posible y deseable pensar las piezas que nos permitirían conformar un modelo alternativo; uno desde el cual se haría plausible la idea de que sí cabe hablar de una evaluación de los mencionados compromisos éticos y valorativos. Con ello aspiro a mostrar de qué modo puede argumentarse en pro de la posibilidad de experiencias morales y modos de ejercer la agencia humana apoyados en un uso menos restringido de la racionalidad práctica; modos más complejos que los permitidos por el modelo que se obtiene de discursos como el de Gauthier. Me refiero a formas de experiencia moral, formas de ejercer la agencia moral y formas de deliberar y decidir a las que creo que resultaría difícil excluir de nuestra experiencia como agentes humanos; a las que tal vez no quisiéramos renunciar, ni tampoco considerar como meras expresiones de nuestra irracionalidad, o de aquello que preferiríamos no tener que evaluar o justificar ni ante otros agentes, ni ante nosotros mismos.

A continuación, en este primer capítulo dedicado a Gauthier se busca identificar las tesis de fondo, la estrategia argumentativa general y los cimientos de su propuesta. En el segundo capítulo se expondrá su teoría de la justicia, importante dentro del esquema de su *Morals by Agreement*, dado que con ella se busca dotar a dicho esquema del elemento que muchos de sus críticos echarían en falta: un contenido propiamente moral. Éste podría haberse difuminado a causa de la insistencia de nuestro autor en el mercado, al que postula como modelo de todo sistema de interacciones humanas y, en estrecha conexión con ello, en la figura del *homo oeconomicus*, al que postula como paradigma de agencia moral-racional. Finalmente, en el capítulo tercero se concluye el examen de la propuesta de Gauthier, para lo cual se lleva a cabo un análisis de su intento por hacer del *homo oeconomicus* un personaje menos amoral, mediante su estrategia de mostrarlo más cercano a los entusiasmos y sensibilidades propias de otro personaje fundamental en el discurso del filósofo canadiense: el individuo liberal.

⁵ Tomo este apelativo de J. Habermas (1990), quien, a su vez hace eco de la expresión utilizada por Ch. Taylor (1994).

Capítulo 1

El nexa entre moral y autointerés: los supuestos iniciales de *La moral por acuerdo*

La propuesta de Gauthier tiene como eje su tesis de que existe una conexión esencial que, a la luz de las ya mencionadas reducciones de los conceptos de moralidad y racionalidad, tal vez podría parecerle insólita a un lector desprevenido: la relación que, según el autor, une a la moral con el autointerés, esto es, con la búsqueda que cada quien hace de su propio beneficio. El propósito del filósofo canadiense será incluso más ambicioso: mostrar que las normas morales hacen parte de la teoría contemporánea de la elección racional, esto es, del corpus axiomatizado de los principios que deben guiar las elecciones de agentes interesados en maximizar su utilidad *i.e.* no en restringirla ni en contribuir a la utilidad de otros agentes.

Creo que en principio esta conexión entre restricciones morales y conveniencia individual no tendría por qué parecernos extraña –y hasta podría antojársenos como algo bastante sensato– si por “conveniencia” se entendiera aquello que, en un sentido amplio, “le conviene a” o “es bueno para” un agente racional; y si “moral” no significara solamente limitaciones al autointerés. Pero como ya se ha anunciado, esta ampliación de contenido queda excluida del planteamiento del autor, quien insiste, a todo lo largo de su texto, en los sentidos estrictamente reducidos en que deben entenderse moral y racionalidad, si se quieren evitar compromisos metafísicos o filosóficamente injustificables (1986, pp. 6 y 238). Un importante resultado de esta restricción semántica será, como lo iremos viendo a lo largo de la siguiente exposición que, una vez asumidos los términos de Gauthier, el intento por demostrar esa relación que une, según él, nuestro ser-racionales con nuestro ser-morales, se tornará en un ejercicio bastante complicado.

1. El contrato, la imparcialidad y el rechazo de la moral “de fines”. El tipo de agente que surge de este primer esquema

Estas dificultades con las que se encontrará en su intento por establecer que la moral necesariamente se incluye dentro de la conducta maximizadora, para el autor canadiense no sólo explican en buena parte la relevancia de su propuesta sino que, además, constituyen un obstáculo que puede y debe ser solventado acudiéndose, en primer término, a lo que él considera la clave de esta relación de inclusión –repito, acaso insólita a los ojos de un lector desprevenido– entre la moralidad y una racionalidad que está al servicio del autointerés y de la eficiencia técnica: el modelo *contractualista* sobre el que ha de fundarse la moral. En este orden, Gauthier continúa la tradición moderna y contemporánea del contrato social pero, en su caso, el sentido en que se asume la racionalidad es claramente el que propone la teoría de la elección racional. Lo cual marca un significativo contraste con otros discursos que también reeditan la idea del contrato pero asociándolo a concepciones, en mi opinión más amplias, de racionalidad práctica v.g. el neocontractualismo de J. Rawls.⁶ A diferencia de lo afirmado por este último, para el filósofo canadiense las relaciones entre contractualismo y *Rational Choice* son las que explican el vínculo entre ser moral y ser un maximizador autointeresado. La razón de este nexo que cree hallar Gauthier está en que el contrato social expresa, según él, la característica imprescindible que debe tener un sistema de normas morales para que se lo pueda considerar como racionalmente justificado: la *imparcialidad*.

De allí que nuestro autor intente mostrar que los mandatos morales se justifican racionalmente como restricciones al interés egoísta, restricciones que se hacen necesarias precisamente para salvar dicho interés. Pues según el filósofo canadiense, con la moral se salvan los intereses de todos, pero no de un “todos” entendido como un “nosotros” o un colectivo social, sino como una agregación de individuos a quienes se les trata como tales *individuos*; nunca como meras partes de un todo. Así, Gauthier cree

⁶ Véase en Rawls: los conceptos de lo “racional” y lo “razonable”, especialmente en los párrafos 63 y 64 de *Teoría de la justicia*, así como su idea de la relación entre la justicia y la moderna teoría de la elección racional, en los capítulos 2 y 3 de *Justicia como equidad*, y en párrafo 1 de la Conferencia II del *Liberalismo político*. En este último (p. 69) se refiere de modo específico, y de manera crítica, al intento de Gauthier de deducir lo razonable a partir de lo racional: “. . . no es posible pensar en derivar lo razonable de lo racional. En la historia del pensamiento moral algunos lo han intentado”. Con relación a este tema rawlsiano, Cf. también: Freeman (2007a), Pogge (2007), Scanlon (2003) y Larmore (2003).

que en su propuesta los intereses de *cada uno* de esos individuos son tenidos en cuenta, ya que no se le da prioridad (ni se le resta importancia) a los fines de nadie en particular.⁷ En breve se examinará con más detalle esta tesis de que la *imparcialidad* de los principios morales es aquello que permite incluirlos dentro del conjunto de las reglas de *elección racional*. Si los mandatos morales se justifican porque son racionales *i.e.* favorecen el interés de cada uno, esto implica según el autor, que deben rechazarse las posiciones emotivistas⁸ que típicamente suelen asociarse (y creo que con toda razón) a la idea de que la racionalidad práctica es razón técnica-instrumental o solamente razón “de medios”, no razón “de fines”. Para Gauthier, en contra de las tesis que le atribuye al emotivismo, sólo si se asume una concepción de la razón como exclusivamente técnica-instrumental, concepción que suscribe fervorosamente el autor de *La moral por acuerdo*, es como puede mostrarse que las reglas morales son un subconjunto de los principios de elección racional. Por tanto, sólo desde una posición como la suya puede verse claramente, según él, que el cumplimiento de reglas morales hace parte de una conducta maximizadora, no de un altruismo con el cual un agente pueda comprometerse por motivos que escapen a los estándares de dicho tipo de conducta, o que queden por fuera de una justificación racional, en el sentido de una razón a la que se acude únicamente buscando adecuar los medios a los fines. Incluso Gauthier va más allá: intentará mostrar que la moral se explica por razones no-morales, por mor de utilidad, y por ello se justifica, racionalmente hablando, que queramos actuar moralmente.

We claim to generate morality as a set of rational principles of choice. We are committed to show why an individual reasoning from non-moral premises, would accept the constraints of morality on his choices.⁹

⁷ Esta tesis (ya lo habrá advertido el lector) en principio sería compartida por Rawls, a pesar de la distancia crítica que intenta asumir Gauthier frente a éste. Como es sabido, el autor de *Liberalismo político* también centra su idea de la justicia en la imparcialidad o equidad (*fairness*) y en la no-instrumentalización de los agentes que suscriben su versión, también contractualista, del pacto social. Hasta aquí, por tanto, pareciera que la diferencia entre ambos autores reside en que el canadiense insiste en una noción estrictamente *maximizadora* de la racionalidad que caracterizaría a los agentes que suscriben el contrato; diferencia que, sin embargo y como iremos viendo, traerá consecuencias de la mayor importancia. Adicionalmente, Gauthier es enfático al negar que haya, por parte suya, un interés en la equidad, y menos aún, en la igualdad (1986, pp. 217 y ss.).

⁸ Cf. Gauthier (1986, pp. 2, y 55 y ss.).

⁹ Gauthier (1986, p. 5). En lo que sigue de estos tres primeros capítulos, a menos que anuncie lo contrario, todas las citas y referencias que haga dentro del texto se tomarán de *Morals by Agreement*. Anotaré al final de cada cita únicamente el número de página, sin hacer referencia al año de publicación de dicha obra (1986).

De allí que, según nuestro autor, su programa de fundamentación racional de la moral sea más exitoso y más radical que los intentos de otros pensadores, como J. Harsanyi, quien ha buscado mostrar que la racionalidad moral hace parte de la racionalidad económica.¹⁰ Según el filósofo canadiense, Harsanyi afirma que la elección de actuar bajo el utilitarismo de la regla (*rule utilitarianism*) se hace, en último término, por mor de moralidad o porque el agente quiere ser altruista. En contraste con ello, Gauthier sostiene que los principios morales vienen exigidos por mor de racionalidad; esto es, se quiere ser moral en aras de una mayor utilidad y, de este modo, se elige no ser egoísta como parte de una estrategia racional que *no* se explica por motivos altruistas. Si para Harsanyi, según nuestro autor, no existe algo así como un requerimiento racional de ser morales, en su caso, por el contrario, este requerimiento puede ser demostrado si se logra mostrar a su vez, y este es el objetivo de Gauthier, que existen unos principios de elección racional que son *imparciales* y que no son otros que los principios de la moral. Entendiéndose por esta última algo que no necesariamente tiene por qué coincidir con las ideas tradicionales o pre-filosóficas que tengamos acerca de la moralidad¹¹.

We claim to demonstrate that there are rational constraints, and that these constraints are impartial. We then identify morality with these demonstrated constraints, but whether their content corresponds to that of conventional moral principles is a further question, which we shall not examine in detail. No doubt there will be differences, perhaps significant, between the impartial and rational constraints supported by our argument, and the morality learned from parents and peers . . . But our concern is to validate the conception of morality as a set of rational impartial constraints on the pursuit of individual interest, no to defend any particular moral code (p. 6).

Como ya se ha dicho, nuestro autor manifiesta insistentemente su compromiso con la idea de que la moral filosófica o correctamente entendida no es otra cosa que el conjunto de restricciones a la conducta egoísta. Este compromiso, según Gauthier, es consistente tanto con su intención de no defender ninguna idea *sustantiva* de moralidad,

¹⁰ Véase la crítica de Gauthier a J. Harsanyi (Gauthier, pp. 3-6, 8, 24, 38, 44, 69-70, 75, 79, 85, 126-127, 146, 149-150, 199-200, 235, 238-245); y véase Harsanyi (1982; 1983 y 1985).

¹¹ En esta oposición entre moral-filosófica y moral-popular, Gauthier coincide con los utilitaristas contemporáneos, como el mismo Harsanyi o como J. J. C. Smart, si bien creo que uno tendría que concederle a Harsanyi que éste, a diferencia de Gauthier y de Smart, si le reconoce —aunque tal vez, muy a su manera— un lugar importante a la moral tradicional. Véase Harsanyi (Op. cit.), y Smart (1988).

como con su fidelidad a una concepción de racionalidad práctica en tanto que racionalidad exclusivamente de *medios*. Pues para el autor de *La moral por acuerdo*, los *fin*es señalan, precisamente, *valores sustantivos*, apuestas morales concretas propias de los códigos morales típicos de aquello a lo que él se refiere como moral “convencional” o moral “tradicional”. Dicha forma pre-filosófica o no-filosófica de moral, a los ojos de Gauthier, es claro que no puede proveer los contenidos de una moral fundamentada racionalmente. Por lo tanto, es tarea del filósofo que emprenda tal programa de fundamentación, el llevar a cabo una sana “purga” de los mandatos morales para que éstos no sean la expresión de compromisos valorativos de carácter sustantivo, los cuales, según nuestro autor, estarían viciados de parcialidad, no pudiendo ser justificados como normas válidas para *todo* agente racional. Creo que en este punto bien podría uno preguntarse si acaso la confianza de Gauthier en su idea de imparcialidad como la piedra de toque de la moral y como gozne que la une a la racionalidad, no sería, a su vez, una apuesta por un valor sustantivo e histórica-culturalmente cargado, que tal vez pueda defenderse exitosamente, pero por otras vías que no son las que utiliza nuestro autor.¹² Sin embargo, pocas veces a lo largo de su texto Gauthier hará eco de esta inquietud que bien puede plantearsele, sobre todo, como se verá en lo que sigue, en aquellos pasajes en los que el filósofo canadiense intenta mostrar la superioridad moral de la sociedad de mercado. Dejo señalado este asunto, sobre el que volveremos reiteradamente en este capítulo.

1.1 El uso de la noción de imparcialidad

En relación con su idea de imparcialidad y el papel tan definitivo que Gauthier parece concederle, creo que también podría uno preguntarse si no sería, o bien un tanto ingenuo, o bien artificioso asignarle a la imparcialidad el peso de ser el vínculo que une a la moral con una racionalidad exclusivamente *maximizadora*, hasta el punto de explicar la inclusión de la primera dentro de la segunda. Pienso que sería fácil estar de acuerdo en que la imparcialidad es una virtud *prima facie* de aquellas reglas morales que estamos dispuestos a aceptar sin reticencias. Máxime tratándose de nosotros, en

¹² Este señalamiento puede hacersele, según Robert Goodin (1993), no sólo a Gauthier sino, en general, a todos aquellos que se inscriben dentro de la tradición del contrato social e, incluso, a algunos utilitaristas, como Harsanyi.

razón del –tal vez doloroso– recorrido histórico que ha marcado la formación de nuestra sensibilidad moral como occidentales modernos; dadas las duras discusiones y los duros aprendizajes que se han vivido en el seno de nuestra cultura y que se ven reflejados en los textos más importantes de nuestros filósofos morales. Nadie que se haga cargo de nuestras intuiciones morales más básicas estaría dispuesto a negar que, para nosotros, la imparcialidad es *conditio sine qua non* de un código moral mínimamente aceptable o defendible en la esfera pública.¹³ Creo que también podríamos aprobar sin reticencias (e incluso como una perogrullada) la afirmación de que las reglas de la elección racional son válidas para *cualquier* agente racional que intente serlo y busque por ello mismo maximizar su utilidad. De allí que dichas reglas valdrían sin distinguos de las características o circunstancias personales de quienes puedan hacer uso de ellas. No obstante pienso que aquí bien puede señalársele a Gauthier el siguiente problema: una vez que un individuo concreto asume de hecho la conducta maximizadora e intenta cumplir con los requerimientos de las reglas de la elección racional, no se ve claro que por ello deba hacer caso omiso de sus circunstancias y, sobre todo, de sus intereses personales. Lo cual sería, por lo demás, racionalmente incorrecto, tanto desde el punto de vista paramétrico, como estratégico.¹⁴

Tal vez puede haber un sentido en el que se podría convenir con Gauthier en que el agente específico que trate de aplicar las reglas de la elección racional estaría intentando que sus elecciones sean racionalmente justificadas y, por tanto, “universalizables” o aceptables para “cualquier” individuo que estuviese en su lugar e intentara maximizar su utilidad. Esto es, trataría de elegir como una suerte de agente “abstracto” que asume el punto de vista *normativo* que tienen los principios de la elección racional para *cualquier* maximizador. Pero el caso es que no pienso que esta actitud pueda ser vista como la de un agente que busca ser “imparcial” en el sentido específicamente *moral* del término, tal y como, en cambio sí lo sería, por ejemplo, un agente kantiano que intentara aplicar en unas circunstancias concretas una máxima que cumpla con los requerimientos normativos del imperativo categórico –universal, imparcial–. Pues estos requerimientos le exigen ante todo –y aquí reside la diferencia

¹³ Me refiero a este aspecto histórico *nuestro* previendo críticas como las que podría hacerme una autora como Agnes Heller, para quien la imparcialidad y la igualdad ante la ley no siempre ni en todos los contextos, han sido la base de los códigos morales. Véase Heller (1990).

¹⁴ Véase las definiciones de racionalidad estratégica y de racionalidad paramétrica en: Resnik (1998).

propiamente *moral* que creo que habría frente al agente maximizador de Gauthier—olvidar sus propios intereses, fines, afectos y circunstancias personales, para considerar sólo la adecuación de la máxima al principio general-imparcial *i.e.* racional en términos kantianos.

Dicho de otro modo, por una parte pienso que habría que concederle a Gauthier que su agente, al tratar de aplicar las reglas de elección racional, las cuales son normativas para “cualquiera” que intente maximizar su utilidad, en este sentido compartiría con un agente kantiano el interés por tratar de aplicar un principio que, por definición, constituye una instancia normativa, siendo así válido para “cualquiera”. Sin embargo, por otra parte creo que es obvio que, a diferencia del kantiano, el maximizador de Gauthier actúa como un agente individual concreto, autointeresado y parcializado hacia sí mismo y hacia sus fines e intereses.¹⁵ La conclusión a la que quiero llevar al lector es que las reglas de elección racional, incluyendo aquellas que se aplican en juegos de cooperación, serían “imparciales” únicamente en un sentido bastante trivial desde el punto de vista *moral*; un sentido que no parece ser el mismo —mucho más fuerte o importante— en que llamamos “imparciales” a las normas morales con las que intenta cumplir un agente kantiano. Por ello pienso que bien puede objetársele a Gauthier que ese nexo que, según él, une a la moralidad con la racionalidad maximizadora, es decir, la imparcialidad, aparece como un concepto bastante difuso que carece, por demás, de sustancia moral o que se pierde en un mero juego de palabras. Tal vez la imparcialidad a la que se refiere nuestro autor no daría tanto de sí, como para constituirse en la base sobre la cual se construya toda una demostración de que las normas *morales* son una subclase de aquellas reglas de elección racional que guiarían la conducta *maximizadora-autointeresada*. En breve se volverá sobre este tema, sobre todo en lo que tiene que ver con el modelo de agente que propone el filósofo canadiense.

¹⁵ Esto sería así, repito, tanto en situaciones donde se imponga el uso de una racionalidad meramente paramétrica, como en aquellas otras en las que se requiera acudir a una de tipo estratégico.

1.2 El agente “no-concernido” y sus razones pre-morales para actuar moralmente. El mercado como modelo de moralidad

Por lo anterior, creo que se hace claro que la preocupación de Gauthier por la neutralidad valorativa y por tomar una distancia crítica y de desconfianza ante la moral tradicional o histórica-culturalmente transmitida estaría, a su vez, enlazada con el modelo de agente moral que puede extraerse de su discurso. Como se verá, es plausible decir que se trata de un individuo que aparece *in abstracto*, concebido sin lazos con otros agentes y sin contexto histórico. De él sólo sabemos que tiene *preferencias* y que es racional cuando trata de satisfacerlas, esto es, que elige buscando *maximizar* su utilidad personal (p. 9). Lo interesante es que, de este agente, el autor destacará un rasgo que a los ojos de alguien ajeno a su discurso creo que puede aparecer, por lo menos en principio, como problemático, sobre todo si se busca fundamentar la *moral*: el agente modélico de Gauthier es un sujeto que *no* está interesado en *los demás*. Mejor aún, los agentes racionales que, según él, se deciden a cooperar entre sí, esto es, que optan por restringir su conducta maximizadora y cumplir con los requerimientos de la moral, lo hacen, afirma nuestro autor, por mero autointerés, no porque estén preocupados por sus congéneres. Son, pues, seres mutuamente “no-concernidos” (*unconcerned*) (p. 17), característica fundamental que obedece a la inspiración básicamente hobbesiana que el filósofo canadiense reconoce repetidamente en su texto, si bien, como veremos, en algunos pasajes intentará tomar distancia frente al autor del *Leviatán*.

En Gauthier, la reedición de la idea del contrato social se realiza, entonces, en clave hobbesiana y en ello creo que encaja, así mismo, otra de sus definiciones: la de “persona”. Como enseguida puede apreciarse, dicha definición no resulta, en mi opinión, fácilmente compatible con la conocida noción rawlsiana de “persona”, entendida como un ser “racional y razonable”.¹⁶ En consonancia con su herencia hobbesiana, la “persona” de Gauthier es un agente racional *i.e.* autointeresado, que arribará al terreno moral “secundariamente” y sólo en razón de sus intereses, los cuales son previos al encuentro del agente con otros individuos y, por ende, también son previos al surgimiento de la moral.

¹⁶ Véase: Rawls, parágrafo 63 de *Teoría de la justicia*. Como podrá verse a partir del capítulo cuarto, habría una distancia aun más llamativa entre la “persona” a la que se refiere Gauthier, y los tipos de agentes a los que, respectivamente, autores como H. Frankfurt y Ch. Taylor llaman “persona”.

A person is conceived as an independent centre of activity, endeavouring to direct his capacities and resources to the fulfilment of his interests. He considers what he can do, but initially draws no distinction between what he may and may not do. How then does he come to acknowledge the distinction? How does a person come to recognize a moral dimension to choice if morality is not initially present? (p. 9).

En este punto creo que el lector podría preguntarse: ¿cómo es que un maximizador autointeresado y no-concernido por otros agentes, un individuo que parte de aquello que nuestro autor supone es el terreno no-moral o pre-moral de sus intereses, puede luego arribar al ámbito moral y hacerlo, paradójicamente, en razón de dichos intereses? La respuesta de Gauthier es que puede mostrarse, y él va a intentar hacerlo, que en ello no hay paradoja alguna, puesto que los principios morales se siguen *luego* de que se establecen los principios de elección racional. De manera que sin asumir ningún compromiso moral previo, sin que se haya partido de algún presupuesto moral, se llega de modo necesario, según el autor, al compromiso con la moral, y ello por razones de utilidad. Dejo de lado, repito, la objeción que podría plantearsele a Gauthier, en el sentido de que un compromiso con la imparcialidad (que, como vimos, es la condición con la cual, según él, deben cumplir los principios de elección racional en contextos de racionalidad estratégica) pueda ser visto como algo distinto a un compromiso *moral*; o que la imparcialidad no sea, de suyo, un *valor* moral. Para nuestro autor claramente este no es el caso, pues según él, los principios morales surgen, en su modelo contractual, de un *acuerdo* que se suscribe entre agentes maximizadores; acuerdo que tiene lugar en un contexto *previo* a la moral misma. De manera que, guiados por una lógica estratégica, los agentes se comprometen a cumplir con las restricciones morales, esto es, llegan a la moral por razones *no-morales*.

. . . [Morality] emerges quite simple from the application of the maximizing conception of rationality to certain structures of interaction. Agreed mutual constraints are the rational response to these structures. Reason overrides the presumption against morality (p. 9).

Según Gauthier esta situación, tal y como es presentada en algunos de los clásicos del contractualismo, concretamente en Hobbes, va ganar estabilidad, ya que en ella los agentes son descritos como dispuestos a seguir cooperando luego de que han suscrito el pacto social, dado que se enfrentan al poder del soberano, que es la instancia

que fuerza a no traicionar dicho pacto.¹⁷ Sin embargo, nuestro autor se separa aquí de Hobbes y dice que mostrará las razones que tienen los agentes para seguir cooperando en ausencia de un soberano, razones que anclan en la necesidad que los individuos tienen de participar, sin sufrir daños, en una estructura de interacción cuyo modelo no es el *estado* de Hobbes, sino el *mercado*. Como se irá viendo en lo que sigue de estos tres capítulos iniciales, esta tesis de la imbricación entre moral y mercado será fundamental en el relato de Gauthier, si bien y como intentaré que salga a la luz, la insistencia en ella creo que le acarreará muchos problemas a la plausibilidad de su propuesta. El autor busca demostrar que si la moral se hace innecesaria en un mercado perfectamente competitivo (en un mercado ideal), por el contrario, en aquellas sociedades con mercados reales, en las cuales éstos pueden presentar fallos¹⁸, se hacen indispensables las restricciones morales, y es por esto que los agentes se comprometen a seguir tales restricciones; no porque se vean sometidos a la coerción ejercida por un soberano. Así según Gauthier, mercado y moralidad se implican mutuamente al instaurar estructuras de cooperación en las que se armonizan, sin necesidad de coerción, los diversos intereses individuales. “Market and morals shares the non-coercive reconciliation of individual interest with mutual benefice” (p. 14).

Por lo anterior, la moral se hace necesaria al permitir que el mercado opere como si se tratase de un mercado ideal o perfectamente competitivo, reemplazando de este modo a la “mano invisible” de Adam Smith, al proveer una armonía artificial entre los intereses de los distintos individuos: cada uno buscando su propio beneficio aportaría, sin buscarlo, al beneficio de todos.¹⁹ O mejor, el aporte que se hace al beneficio *de otros* se hace sin coerción, voluntariamente, como una forma de apostar por el beneficio *propio*. Frente a Hobbes, Gauthier cree que su esquema cuenta, entonces, con la ventaja de la no-coerción, que será indispensable para ganar la aceptación voluntaria, no coaccionada de los agentes y, por esta vía, garantizar la estabilidad y la legitimidad del

¹⁷ Cf. Hobbes (Capítulo XVIII, de la segunda parte del *Leviatán*).

¹⁸ Sobre el clásico tema de los fallos o imperfecciones del mercado y las características que habría que atribuirle a un mercado al que, por el contrario, se lo pudiera pensar como perfectamente competitivo, véase: Bator, Francis M. (1958, pp. 351-379).

¹⁹ Véanse, por ejemplo, las referencias a la “mano invisible”, en la *Investigación sobre la naturaleza y las causas de la riqueza de las naciones*, (trad. 1982, pp. 17 y 402), de A. Smith. Sobre la noción de mercado perfectamente competitivo y su relación con la idea de la mano invisible que parece guiarle, se volverá más adelante.

sistema. Así mismo, y como contrapunto al relato de Smith, a Gauthier le interesa resaltar que, puesto que los agentes para los que él está haciendo su propuesta no participan en un mercado ideal, la búsqueda que cada uno de ellos hace de su propio beneficio necesitará de una serie de restricciones –que serían innecesarias en un mercado sin fallos– a fin de que a nadie se le impongan costes forzosos, y que dichas restricciones beneficien a todos. Por ello, se hace indispensable alcanzar un *acuerdo* racional que cuente con la aquiescencia de todos y en virtud del cual cada uno se comprometa a cooperar. Este acuerdo se dará sólo si cada agente calcula que de dicho pacto podrá obtener para sí mismo el máximo beneficio posible que sea compatible con el máximo beneficio que logran los demás, aportando, a cambio, el mínimo posible de sacrificio. En otros términos, el acuerdo deberá cumplir con el principio básico racional del minimax (en contraste con el maximin de Rawls).²⁰ El reto que Gauthier ve ante sí es mostrar las razones para seguir cumpliendo con el acuerdo *luego* de que éste ha sido suscrito. Para ello propone que se distinga entre dos tipos de agente maximizador: aquel a quien el autor define como un maximizador “restringido” o “constreñido”, y al que me referiré como MR; y otro tipo de maximizador al que Gauthier considera un maximizador “irrestringido”, y al que llamaré MI.

1.3 El maximizador irrestringido MI vs. el cooperador o maximizador constreñido MR. El problema de la motivación moral de este último

El MR es el sujeto dispuesto a cooperar con otros y por tanto, a sacrificar parte de sus beneficios posibles, en aras de garantizar una estructura de cooperación menos riesgosa que, finalmente, también le rendirá sus beneficios. En contraste con esta forma de agencia, está la del MI: el individuo no-cooperador que no está dispuesto a hacer este tipo de concesiones y de allí que, según Gauthier, tampoco podrá acceder a los beneficios que arroje la cooperación.

We distinguish the person who is disposed straightforwardly to maximize her satisfaction, or fulfil her interest, in the particular choices she makes, from the person who is disposed to comply with the mutually advantageous moral constraints, provided she expects similar compliance from others. The later is a constrained maximizer. And

²⁰ Cf. parágrafo 26 de *Teoría de la Justicia*. Sobre los principios del maximin vs. minimax, véase: Resnik (1998, pp. 56-69).

constrained maximizers interacting one with another, enjoy opportunities for cooperation with others lack (p. 15).

La figura del MR será la que asuma el papel de un agente moral, o mejor, un agente moral es, para Gauthier, básicamente un MR. A estas alturas creo que el lector bien puede preguntar: ¿qué explica que este tipo de agente quiera comportarse como un MR y no como un MI, o un maximizador sin más? Adicionalmente, ¿qué garantiza que quiera seguir siendo un MR una vez que ha suscrito el acuerdo en virtud del cual ha renunciado a ser un MI pero que, con posterioridad a éste, se le presenten nuevos incentivos para traicionar dicho pacto? En todo caso, para Gauthier, la garantía de ese cumplimiento *no* puede residir en previos compromisos morales con valores sustantivos, pues esto, amén de viciar de no-racionalidad y de parcialidad el acuerdo previo, resulta ser, a los ojos de nuestro autor, algo bastante incierto e inestable. Las restricciones morales a la conducta maximizadora tienen que contar, según él, con la aquiescencia racional de los agentes, quienes deben “ver por sí mismos” la racionalidad de dichas restricciones. Éstas han de tener un sentido para ellos, un sentido en virtud del cual los individuos se convencen a sí mismos y por buenas razones de que, en su caso, la moral “paga”; llegando a esta conclusión por su propia cuenta, sin que nadie les fuerce ni les manipule para creer irracionalmente en ella. De allí la importancia que reviste para el autor la característica del *no-concernimiento* mutuo, propia de la actitud de cada agente, tal como ya se mencionó.

Las restricciones morales deben ser racionales a los ojos de quien está preocupado por *su propio beneficio*, sin condicionar su motivación para el cumplimiento con ellas a un previo compromiso con el bienestar de los otros; compromiso que se base en valores asumidos pre o extra-racionalmente, tales como los que caracterizan, según Gauthier, a las morales tradicionales o no filosóficas. Nuestro autor estima que su modelo contractualista de moral permite dar cuenta de esa motivación puramente racional y autointeresada para cooperar. Con ello se gana, según él, la ventaja de ofrecer una concepción “débil” y “amplia” de racionalidad práctica, concepción que aporta una base más firme y menos incierta para la moral (pp. 8 y 16). Sólo así podría explicarse de un modo plausible por qué querría alguien comportarse moralmente, sin que para ofrecerse tal explicación tenga que apelarse a una noción más

“fuerte” de racionalidad que incluyera compromisos morales previos. Al contrario de lo que implica esta última estrategia –errónea a los ojos de Gauthier–, para él la moral debe poderse deducir o ha de poder ser justificada como el resultado lógico de la aplicación de principios de elección racional a situaciones de interacción; principios que tendrían que ser axiológicamente neutros, para que puedan motivar/obligar a *cualquier* agente racional (pp. 18-20). Tal vez en este punto se podría objetar que nuestro autor considere su noción de racionalidad como más débil, amplia, neutra y moralmente más vinculante que otras nociones alternativas –y ello, además, suponiendo que estos epítetos necesariamente señalen características positivas, lo cual, en mi opinión, no es evidente–. Por otra parte, creo que sorprende que Gauthier no se pregunte, por lo menos no en la primera parte de su texto, si una motivación no-altruista y no-arraigada en compromisos con valores morales previos a un acuerdo de mutuo beneficio, ¿acaso no sería tan débil que pondría en peligro la lealtad hacia dicho acuerdo? A lo mejor el lector podría cuestionar: alguien que se comprometa a restringir su conducta maximizadora únicamente en virtud de sus expectativas de beneficio, ¿tal vez no sería, precisamente por ello, visto por sus congéneres como un sujeto de poco fiar, como un ser al que difícilmente se le podría llamar “agente moral”? En suma, la cooperación con este tipo de individuo ¿no sería, a lo mejor, percibida como riesgosa por parte de los demás agentes?

Al final de su libro Gauthier reconocerá estas limitaciones de su maximizador MR y tratará de mostrar que éste se parece menos al *homo oeconomicus* y más al “individuo liberal”, figura a la cual el autor intentará presentar como caracterizada por una mayor sensibilidad moral. Con estos ajustes posteriores que Gauthier le hará a su agente modélico, veremos que el autor asume que se solventan estos problemas de falta de motivación y de compromiso propiamente moral, de los que puede acusarse al MR. En su momento podremos comprobar si resulta convincente la solución que nuestro filósofo propone para estas dificultades que él mismo crea. Pero las crea, en mi opinión, y es lo que por ahora quisiera destacar, al haber atado desde un principio su modelo de agente a la estructura del *mercado*, o al haber postulado al mercado como modelo de interacción humana y, con ello, al *homo oeconomicus* como modelo de agente moral. Sobre este problema, que creo puede ser visto como el talón de Aquiles de la propuesta de Gauthier, volveremos de manera reiterada a lo largo de estos tres capítulos, dado que

resulta determinante para el tema del modelo de agencia moral y del modelo de racionalidad asociado a dicha agencia.

1.4 El modelo mecanicista, la moral como anomalía y la moderna sociedad de mercado como “caso ejemplar”

Por lo anterior, a continuación me permito abrir un breve paréntesis, con el fin de señalar algo que considero importante para contextualizar e interpretar las tesis de Gauthier. Algunos comentaristas lanzan serias sombras de duda relacionadas con este problema específico de la amoralidad y la falta de una motivación propiamente moral de la que puede ser acusado el agente MR, así como con el carácter general de la propuesta del filósofo canadiense. Desde distintos puntos de partida, R. Brandom y A. Ripstein señalan que en *La moral por acuerdo* surgen las dificultades propias de una ya larga tradición, cuyas raíces pueden rastrearse en la revolución científica que protagonizó el nacimiento del mecanicismo clásico.²¹ Ambos autores afirman que en un mundo pensado bajo el paradigma mecanicista no habría cabida para los valores ni las normas y, por ello, dar razón de una realidad social en la cual dichos valores y normas parecen seguir operando, exigiría explicarlos como algo que los agentes “introducen” en el mundo con posterioridad a su mutuo encuentro en el espacio social. En otros términos, si se toma demasiado en serio la imagen del mundo tal y como surge bajo este paradigma y se extrapola la realidad física a la social, los valores y las normas que parecen operar en esta última se hacen anómalos.

Por ello no se los podría explicar, a menos que se los postulara como artificios que se agregan al mundo, adicionándoselos a la “verdadera” realidad –física– que les antecede y que no está hecha de suyo en clave social ni humana. De allí la vieja idea, aún tan presente en algunas propuestas filosóficas del siglo XX, v.g. el positivismo lógico y algunos autores contemporáneos que simpatizan con los supuestos de este tipo de filosofías (como es el caso del mismo Gauthier), de que la moral, tal y como se aprecia en las sociedades humanas que conocemos, debe poder ser “domesticada”; tiene que ser “purgada” y reducida a cálculo racional ya que, de lo contrario, no podría darse

²¹ Véase Brandom (2001) y Ripstein (2001). También resultan muy esclarecedoras las críticas que, señalando problemas bastante cercanos a los que indican estos dos autores, pueden hallarse en algunos textos de A. Wellmer (1988); H. Putnam (2002); y Ch. Taylor (1985d; y 1997, introducción).

cuenta de ella ni de su fundamento; no se vería de dónde viene ni qué sentido tiene.²² Para que la moral y, en general, las normas que parecen guiar las interacciones humanas, no sean una anomalía inexplicable dentro de un mundo pensado mecanicistamente, acaba por considerárselas como un epifenómeno de ese mundo físico “verdadero” o “de base” que antecede a lo humano y lo social. A la moral se la ha de mostrar, entonces, como un resultado posterior que brota del cálculo racional de los individuos. A su vez, la constitución de estos individuos es pensada como una autoconstitución que antecede y moldea a lo social. Así, de la mano de la mencionada extrapolación del modelo mecanicista a lo social, vendría otra imagen conexas a ésta y que se halla muy presente en pensadores como Gauthier, herederos de esta tradición criticada por Ripstein y Brandom: la idea de unos agentes humanos que desde siempre han sido individuos propiamente tales.

Se trata de sujetos autoconstituidos, que desde siempre cuentan con una plena capacidad de agencia y que son el origen de la sociedad. Nunca al revés: la sociedad no puede anteceder ni moldear a los individuos; no incide, por tanto, en su capacidad de agencia. Ellos no necesitan la sociedad para ser lo que son, es decir, para constituirse como tales sujetos-agentes, mientras que, por el contrario, lo social sólo se puede explicar como el resultado de la agregación de los individuos y el acuerdo o la negociación entre ellos. De allí que dar cuenta de cómo surge la moral –que ahora aparece como un extraño nexo entre los agentes humanos– equivalga a dar razón de una auténtica anomalía, la cual se soluciona mediante una estrategia reduccionista.²³

²² Esto mismo se aprecia en algunos autores utilitaristas contemporáneos que, como Smart (1988) y Harsanyi (1982, 1983 y 1985), aplican de manera renovada los intentos de sus antecesores, los utilitaristas clásicos, por dar cuenta de la moral a partir de claves mecanicistas. Otro tanto puede atribuírsele al intento de los expositores más visibles del emotivismo contemporáneo, Mackie (1990) y Stevenson (1984). En el quinto capítulo se verá la crítica de Ch. Taylor a este intento reduccionista que, según él, es propio de posturas como el utilitarismo, pero también de algunos programas de investigación de las ciencias humanas a partir del siglo XX, v.g. el conductismo psicológico, que intentan trasplantar el modelo explicativo de las ciencias “duras” a los discursos acerca del hombre. Este intento requiere, como pieza fundamental, la transformación del lenguaje cotidiano que se utiliza para hablar de los eventos humanos, con el fin de que dicho lenguaje sea “neutro” al referirse a tales “hechos”, para lo cual tendría que ser un lenguaje “purgado” de términos valorativamente cargados.

²³ Dicha estrategia reduccionista, que señalan, repito, cada uno a su manera y desde distintos ángulos, autores como Brandom y Ripstein para el caso de Gauthier (y, en general, tratándose de algunos programas filosóficos del siglo XX, tales como el positivismo lógico), también ha sido objeto de análisis, como arriba lo comento en un pie de página, por pensadores provenientes de diversas escuelas, v.g. Ch. Taylor (1985d, introducción, pp. 1-12;), H. Putnam (2002), A. MacIntyre (1987) y A. Wellmer (1988). Entre otras facetas de la cuestión, estos autores muestran algunos de los resultados que traería el hecho de que una importante corriente de la tradición filosófica moderna y contemporánea aplique esta solución

Esta necesidad de introducir la moral en un mundo que le es ajeno, supone ingeniarse algo que en mi opinión admite ser visto, a su vez, como verdaderos artificios que permitirían explicarla a partir de ese mundo no-moral. Dentro de esos artificios hay uno que acaso resulte especialmente revelador: la ausencia de contextos históricos-culturales dentro de los cuales se insertarían los individuos y los códigos morales que parecen guiarles o que, por el contrario, puedan ser objeto de crítica por parte los primeros. Así, se señala un origen ahistórico, no-concreto sino puramente conceptual de lo moral como algo neutro, válido para cualquier agente racional en cualquier tiempo y lugar.

En Gauthier, esta neutralidad y esta ausencia de contexto de lo racional y de lo moral, pienso que sufren una vuelta de tuerca muy interesante: el autor no sólo asume sus nociones de lo racional y de lo moral como neutras, sin hacer caso de la sospecha de que éstas puedan estar cargadas valorativa e históricamente –cosa que, por lo demás, creo que no tiene por qué ser vista como necesariamente negativa o vergonzosa–. Lo extraño es que, amén de lo anterior, de tanto en tanto Gauthier parece que se viera en la necesidad de hacer unas curiosas apologías de la *moderna sociedad occidental* y, sobre todo, de la moderna sociedad de *mercado*. Lo cual equivaldría a defender, paradójicamente, una forma histórica-concreta de sociedad y, por tanto, traicionar su propia consigna de neutralidad. En estas apologías nuestro autor intenta demostrar que este tipo concreto (repito, históricamente datable) de sociedad es moralmente superior a otros, y lo es –con lo cual surgiría otra paradoja– por gozar de una especial neutralidad en virtud de la cual la moralidad y la racionalidad que Gauthier le atribuye a la moderna sociedad de mercado, son válidas para todo tiempo y lugar. Es como si en dicha sociedad se realizara (aunque el autor no lo diga en estos términos) el ideal hegeliano de la realidad que es racional.²⁴ De allí que Gauthier afirme en muchos apartes de su *Morals by Agreement*, que en la moderna sociedad occidental y, sobre todo, en la institución del *mercado* se encarna el ideal de cooperación social que él propone con su “moral por acuerdo”. Pues, según él, en nuestras modernas sociedades occidentales sí hemos hallado la clave para interactuar de tal manera que cada uno, buscando su propio beneficio, aporte al mismo tiempo al beneficio de todos (pp. 101-102, 231, 289-298).

reduccionista no sólo a temas morales y políticos, sino también a ciertos desarrollos de las ciencias sociales.

²⁴ Cf. Ovejero, Félix, (1994a y 1994b).

Esto explicaría la ejemplaridad y el valor normativo que el autor atribuye a nuestra sociedad, en claro contraste, según él, con otras sociedades previas o contemporáneas a ella. Señalo este énfasis del discurso de Gauthier no porque malévolamente insista en mostrar sus flancos más débiles, sino porque creo que resulta ser bastante revelador del modelo de agente moral y de racionalidad práctica que puede extraerse de su texto. Centrémonos, pues, en dicha noción de racionalidad, para lo cual a continuación se hará un breve repaso de algunos conceptos básicos de la teoría de la elección racional y se intentará mostrar hasta qué punto nuestro autor se ciñe a ésta.

2. El modelo de la teoría de la elección racional y la reforma normativa que propone Gauthier

Para llevar a cabo su justificación racional de la moral y con ello desarrollar su propuesta de una “moral por acuerdo”, Gauthier parte, como ya se ha mencionado, de aquella noción de racionalidad que en su momento fue formulada de manera ejemplar por Hume y que contemporáneamente tiene su mejor expresión en la teoría de la elección racional. Dicha noción es la que nuestro autor considera la más acertada, la más clara y mejor desarrollada, hasta el punto de ofrecer, para el filósofo canadiense, el modelo que han de tomar como supuesto imprescindible las ciencias sociales y, en general, todo intento de dar cuenta de las acciones humanas. Veamos, entonces, la versión que ofrece Gauthier del esquema básico de la teoría de la decisión, lo que hereda gustoso de ésta y aquello que, por el contrario, piensa que debería ser objeto de ajustes para mejorar la teoría.

2.1 El esquema de la teoría de la decisión

Siguiendo las nociones iniciales de la teoría de la elección racional, para Gauthier las acciones humanas pueden ser vistas como el producto de unas *elecciones* –elecciones de *acciones*– que se explican o se prevén causalmente a partir de las *preferencias* de los agentes. A éstos se les considera “racionales” en tanto se supone que intentan *maximizar* la medida de dichas preferencias, esto es, su *utilidad*.²⁵ Al mismo

²⁵ Cf. Sen, Amartya (1992).

tiempo, las preferencias y las acciones llevadas a cabo para satisfacerlas se explican por los *deseos y creencias* del agente, de acuerdo con el clásico esquema de la moderna teoría de la acción: el agente A lleva a cabo la acción B porque A tiene el deseo X y la creencia Y de que B es un buen medio para satisfacer X.²⁶ Gauthier parte de un supuesto del que, según él, también han partido ciertos desarrollos de teoría de la decisión²⁷: una sana medida de asepsia metafísica y neutralidad valorativa, desde la cual se exigiría omitir cuestiones tales como la *naturaleza* de las creencias y los deseos de agentes concretos, puesto que dichos deseos y creencias no serían objeto de una observación o de un control empírico directo. Para ofrecer los criterios de racionalidad de las elecciones humanas en general bastaría, entonces, con un presupuesto ontológico muy débil: que dichos deseos y creencias, cualesquiera que sean, pueden ser asumidos como aquello que explicaría las preferencias de cualquier agente racional, esto es, uno que intente satisfacer esas mismas preferencias mediante la elección de sus acciones.

Como puede adivinarse, este supuesto se encuentra en tensión con el interés que podría haber en que los principios formulados por la teoría, así como los modelos formales que desde ella se propongan, sean aplicables, en el sentido de que aunque puedan ser expresados formalmente, también se logre hacer de ellos una herramienta útil para el economista, el filósofo o el científico social que quiera describir, explicar y prever las elecciones humanas, así como responder (a) –o incidir (en)– las preferencias que explican dichas elecciones. Para lograr esta finalidad práctica sin traicionar la consigna de asepsia metafísica y neutralidad valorativa, puede acudir a una salida indirecta: utilizar la información proporcionada, no por unos deseos y creencias que no pueden observarse directamente, ni siquiera por aquello que dichos deseos y creencias se supone que producen inmediatamente, es decir, las preferencias del agente tal y como él las puede experimentar *in foro interno*, sino la *conducta observable de elección* con la que se supone que el agente buscaría satisfacer dichas preferencias. Esto es, se

²⁶ Cf. Davidson, Donald, (2001). Salvo mi paráfrasis y mi traducción.

²⁷ En este aparte se dará cuenta de la versión que presenta Gauthier de la teoría de la elección racional, a saber, aquella que algunos comentaristas llamarían “normativa”, la cual estaría interesada en el aspecto teórico y formal de las elecciones que llevarían a cabo agentes ideales. Esto contrasta con otros desarrollos de la teoría, los cuales podrían caracterizarse como más “descriptivos”, puesto que se centran en el estudio de las elecciones llevadas a cabo por agentes reales, con el fin de incorporar este conocimiento a una teoría que también tenga valor prescriptivo. Esta diferencia no tiene que ser asumida de modo tajante, si bien, como en breve se verá, en el caso específico de Gauthier parece ser tomada bastante en serio. Cf. Resnik, M. (1998).

utilizaría la información proporcionada por lo que el observador alcance a “ver” de la conducta de elección que manifiestan los agentes y, sin necesidad de lanzar tesis ontológicas ni juicios de valor, se pueden *deducir las preferencias* de los agentes a partir de dicha conducta. En otros términos, se partiría de sus preferencias *reveladas*²⁸: aquellas que se deducen de las elecciones de un agente, elecciones que sí pueden ser objeto de observación y, eventualmente, de previsión.

De este modo, se toma en consideración lo que puede hacer un observador que simplemente “ve” que en determinadas situaciones de elección el individuo A suele simpatizar más con el resultado o el estado de cosas X, que con el resultado o estado de cosas Y. Es decir, el observador puede concluir, a partir de la conducta desplegada por A, que A manifiesta una cierta *preferencia* por X antes que por Y. Ahora bien ¿cómo medir *qué tanto* es que A prefiere X a Y, y cómo dar cuenta de dicha preferencia y de su relación con las demás preferencias de A? Aquí habría que hacer uso de otro supuesto: dicha medida vendría dada por el *mayor peso* que tendría para A su deseo de X si se lo compara con el peso que tiene su deseo de Y. Pero esto último, si se tiene en cuenta lo dicho anteriormente con respecto a la asepsia valorativa y metafísica, también envolvería dificultades puesto que un término como “deseo”, o expresiones parecidas, que podrían ser utilizadas para designar aquello que explicaría que A suele preferir X a Y, expresiones tales como “atracción”, “necesidad”, “inclinación”, etc. no designarían objetos directamente observables. Si antes se ha dicho que habría que *deducir* las preferencias de los agentes a partir de su conducta observable de elección, entonces ahora, de acuerdo con lo anterior, habría que también “deducir”, sin incurrir en falta de rigor, una cuantificación adecuada o una expresión cuantitativa o numérica adecuada del “mayor peso” o la mayor “fuerza” que tienen unos deseos de A, comparada con el peso o la fuerza que tienen otros de sus deseos.

Dicha cuantificación se hace necesaria si se va a efectuar algún tipo de cálculo o se intenta llegar a alguna forma de predicción. De modo que habría que ofrecer una expresión numérica o una cuantificación, a pesar, repito, de la dificultad que entraña el hecho de que tal fuerza/peso tampoco es directamente observable ni, por tanto, mensurable. Por tal razón se requiere de nuevo alguna salida indirecta que, para el caso

²⁸ Gauthier pp. 22 y ss. Sobre la noción de “preferencia revelada”, Cf. Resnik (1998), así como la crítica de A. Sen (1992).

de la medición de ese peso/fuerza, implique dar cuenta, no directamente de tal magnitud, sino de lo que pueda observarse y pueda atribuírsele como su efecto en la conducta observable de elección. La solución será acudir a una medida estadística: aquella que vendría proporcionada por la mayor *probabilidad* de que en aquellas situaciones en las cuales A pueda elegir entre X e Y, elija la primera antes que la segunda opción. Dicha medida probabilística/estadística es designada, precisamente, mediante un término clave: “utilidad”. Este vocablo tendría la ventaja de que, amén de que nos evita caer en los problemas presentados por expresiones tales como “deseo”, también podría designar la medida de aquello que intenta maximizar un agente racional con sus elecciones, las cuales, como se ha dicho, serían la expresión de sus preferencias. En conclusión, A es considerado un agente “racional” en tanto que busca *maximizar su utilidad*, lo cual significa que en situaciones en las que se le presente la oportunidad de escoger entre X e Y, sabremos que efectivamente ocurre que A prefiere X a Y porque en dichas circunstancias el resultado más probable es que A termine por elegir X a Y. Las elecciones de A son racionales si se corresponden con sus preferencias, es decir, si A elige buscando maximizar su utilidad esperada. Esto puede ser deducido por el observador si éste advierte cierta coherencia en la conducta de elección de A. Así, un individuo es considerado “racional” en tanto que a sus elecciones se las puede calificar de coherentes, deduciéndose de ellas un sistema –también consistente– de preferencias y, por tanto, si puede afirmarse, en vista de su conducta recurrente de elección, que el agente busca satisfacer sus preferencias *i.e.* maximizar su utilidad esperada. En otros términos, racionalidad es, justamente, maximización de utilidad.

Hasta acá, Gauthier suscribe estas nociones básicas de la teoría de la elección racional aunque, como veremos, también se propone introducirle algunos cambios. Pero antes de continuar, en este punto me atrevo a señalar la coherencia que puede apreciarse entre la versión que hace nuestro autor de aquellos aspectos en los que dice coincidir con la teoría de la decisión y lo que arriba se comentó con respecto a cierto sesgo reduccionista en la propuesta de Gauthier, sesgo que en este aparte se vería reflejado en la asepsia metafísica y la neutralidad valorativa de los supuestos que el filósofo canadiense comparte con ciertas versiones de la teoría de elección racional. Creo que en Gauthier la aplicación de estos supuestos cuando se busca dar cuenta de los criterios para la racionalidad/irracionalidad de las decisiones y acciones humanas se traduce en

un conductismo de fondo²⁹, desde el que se evita apelar a referentes que vayan más allá de la mera conducta observable de elección, referentes que tal vez permitirían hacer menos opacas nociones tales como “preferencia” y “racionalidad”. De hecho, pienso que este conductismo de fondo puede verse en la utilización que hace nuestro autor, tanto del concepto de preferencia revelada como, en general, del esquema y los supuestos que le atribuye a la teoría de la elección racional. En todo lo cual cabría advertir otro de los parentescos reduccionistas del filósofo canadiense, parentesco que nos lleva a recordar la crítica, arriba comentada, que Brandom y Ripstein hacen de la propuesta de Gauthier. En breve se verá cómo dicho parentesco se puede apreciar también —e incluso a pesar suyo— en la centralidad que sigue teniendo, en el discurso del canadiense, la noción de *preferencia* tal y como la hereda de la teoría de la elección racional.

No obstante lo anterior, en algunos aspectos nuestro autor expresamente dice que intentará distanciarse de ciertos aspectos de dicha teoría, concretamente, del hecho de que en ella puede acusarse la ausencia de un elemento *normativo* que vaya más allá de la mera *consistencia* lógica de los sistemas de preferencias. Gauthier considera que dicho elemento se hace necesario si se quiere dar cuenta de qué es lo que permite que tales sistemas sean evaluados como racionales/irracionales, en lo cual creo que no le falta razón. Empero, igualmente pienso que esta ausencia de lo normativo no debería sorprender, dado que como ya se habrá hecho patente con lo dicho en lo que va de este aparte, los referentes normativos desaparecen si se iguala elección a conducta racional, y conducta racional a manifestación de preferencias. Si cualquier conducta de elección del agente A resulta siempre explicable —y sin necesidad de acudir a una complicada teoría de la mente— como la expresión de las preferencias de A, esto es, como la

²⁹ Para una crítica al conductismo lógico, véase el ejemplo clásico de la réplica de J. Fodor (1985) a G. Ryle (2002). Como puede apreciarse en dicha réplica, la idea de Ryle de que la psicología debería referirse únicamente a conductas observables, resulta muy cercana a ciertas posturas propias de la contemporánea filosofía de la mente, las cuales, a su vez, hunden sus raíces en la tradición heredada del programa positivista —v.g. el eliminacionismo, el programa reduccionista de eliminar los términos psicológicos, o la propuesta de “traducirlos” a términos que designen eventos neuronales—. Creo que dichas propuestas admiten ser vistas como parte de una estrategia reduccionista más general, que incluye de manera importante la consigna de “purgar” el lenguaje de las así llamadas “ciencias especiales” (v.g. la psicología, la sociología), con el fin de que éstas sólo utilicen términos observacionales proporcionados por las “ciencias básicas” (v.g. la física, la química, la neurociencia). Aquí puede advertirse otro de los parentescos que pienso que puede atribuirse a la idea de Gauthier y de otros autores, de que la moral, como arriba se señaló, pueda ser explicada en términos que designen realidades a las que se considera más “básicas” que ella misma, y de las cuales ella sólo es un epifenómeno.

maximización de su utilidad esperada, entonces, para hacer un *juicio evaluativo* que busque establecer qué tan “racionales” sean la conducta de elección o las preferencias de A, bastaría simplemente con un único criterio normativo, y bastante débil, que permitiera evaluarlas en este sentido: la *consistencia* lógica del sistema de preferencias de A. Es importante señalar que aquí se habla de *sistema* pues, como veremos, Gauthier también advierte que *no* se puede/debe evaluar la racionalidad de *una* preferencia considerada aisladamente. Así, cualquier conducta de elección y cualquier sistema de preferencias pueden ser considerados “racionales” sólo a condición de ser consistentes,³⁰ y ninguna preferencia admitiría ser vista como racional ni como irracional de suyo, sin que previamente se hayan examinado sus relaciones con el sistema de preferencias al que pertenece.

En esto último pues, Gauthier estaría de acuerdo con la teoría de la elección racional, pero se separa de ésta en lo que se refiere al vacío normativo que advierte en ella, suscribiendo, por ende, algunas de las críticas más importantes que dicha teoría ha recibido. Nuestro autor admite que las mencionadas críticas aciertan al señalar que el mandato “maximiza tu utilidad” dice muy poco sobre cómo ser más racional. Entre otras cosas, porque tiene la simpleza de las tautologías; equivaldría a ordenar algo parecido a “haz lo que quieras (hacer)”.³¹ Para superar esta falencia de la teoría, Gauthier se propone complementarla con la introducción de criterios normativos que permitan evaluar algo más importante que la mera consistencia lógica de los sistemas de preferencias de los agentes humanos. Dichos criterios permitirían *sensu stricto* establecer si estas últimas, así como las conductas de elección de las que hacen parte, son o no propiamente “racionales”.³²

³⁰ Es decir, en tanto cumplan con principios formales, tales como el de transitividad, principios que han sido expuestos y desarrollados en los textos clásicos de teoría de la decisión, como los de Von Neumann y Morgenstern (1953); Luce y Raiffa (1958).

³¹ Véase, por ejemplo, las críticas de autores como N. Rescher (1993); A. Sen (1992), y R. Nozick (1995).

³² Un poco en consonancia con la advertencia que he hecho en un pie de página anterior, habría que señalar una característica muy importante de la noción de “racionalidad”: ésta no sólo ofrece un aspecto *descriptivo* —v.g. explicar la “lógica” de un mecanismo, de un modo de operar, de una acción, de una institución, de un sistema de incentivos, o de un esquema deliberativo—. También se trata de una noción que admite que se la asuma con un interés *normativo* y valorativo, usándose el apelativo “racional” para *elogiar* o calificar positivamente a los agentes, a sus decisiones, a las instituciones y diversas formas de praxis social. Véase al respecto: A. Wellmer (1988), y F. Broncano (1988).

2.2 Las preferencias cualificadas o “consideradas” y el problema de lo normativo

Por lo anterior, Gauthier se propone sugerir algunas mejoras al andamiaje conceptual básico de la teoría de la decisión para que, de ahora en adelante, ésta cuente con el criterio normativo que le hace falta. Tal criterio puede ser ofrecido si se repara en la diferencia *cualitativa* que, según el autor, distingue a un tipo especial de preferencias: aquellas que él propone llamar (p. 27) preferencias “consideradas” (*considered*). Creo que nuestro filósofo introduce esta noción en vista de los aprietos en los que se vería una vez que ha decidido comprometerse con el modelo de racionalidad que le atribuye a la teoría de la elección racional. Pues estaría claro que el precio que terminaría pagando por este compromiso, sería la renuncia a toda posible *evaluación de las preferencias*, esto es, a cualquier apelación a lo normativo, ya que toda instancia normativa desde la cual se intente llevar a cabo tal evaluación, colapsaría, finalmente, en la mera expresión de preferencias no susceptibles de crítica. Como puede preverse, este sería un resultado indeseable para una propuesta *moral* como la que pretende hacer Gauthier. Pienso, entonces, que el autor utiliza su noción de preferencias “consideradas” para evitar dicho resultado y proporcionarle así un peso normativo, tanto a la noción misma de preferencia —que, visto lo anterior, pareciera estar blindada frente a cualquier intento de evaluación—, como al esquema general que el canadiense ha presentado de la teoría de la decisión y, por supuesto, también al modelo de racionalidad que se desprende de éste; en síntesis, tres de los compromisos que parecen ser irrenunciables para nuestro filósofo.

Gauthier reconoce que muchas de las preferencias que tiene cualquier agente no han pasado por un proceso de evaluación, reflexión o cualificación por parte de éste y simplemente se manifiestan en su conducta de elección, a veces sin que el propio sujeto sea conciente de ello. Estas preferencias son, pues, algo muy parecido a las preferencias “reveladas” a las que se refieren, según el autor, los teóricos de la elección racional. Pero también puede ocurrir que un agente tenga otro tipo de preferencias que sí han pasado por un proceso de examen, y con respecto de las cuales el individuo expresa un compromiso en el largo plazo. Se trata, precisamente, de esas preferencias que Gauthier propone llamar “consideradas”. Según él, éstas han sido formuladas por el sujeto bajo ciertas condiciones que las cualifican racionalmente. Por ejemplo, el agente las hace

suyas contando con la información relevante; se fundan en creencias plausibles o adecuadas; no son el producto de la ignorancia o de estados psicológicos adversos, como el miedo o la sugestión; no son el mero producto de la manipulación, ni de los traumas de infancia, ni de pasiones que obnubilen el juicio, como el odio irracional, y un largo etcétera.³³

Según Gauthier, para que la formulación de estas preferencias consideradas se haga posible, es necesario que el individuo las haga *concientes* y, por ende, que también las *verbalice*, las exprese en forma de proposiciones. Pues sólo así puede examinarlas y determinar si son o no consistentes con otras preferencias suyas, de manera tal que pueda decidir si se compromete (o no) con ellas. Este compromiso se hace manifiesto en lo que el agente *dice* expresamente preferir, es decir, a nivel de su *discurso* y no sólo a nivel de su *conducta observable* de elección, como en cambio sí ocurre en el caso de las preferencias simplemente reveladas. Todo lo anterior explica por qué, en comparación con estas últimas, las preferencias consideradas son más estables a lo largo del tiempo. Para dar satisfacción a sus preferencias consideradas, a las que el autor también se refiere como “actitudinales” o relativas a las actitudes del agente (*related to attitudes*), éste tendría que maximizar lo que Gauthier llama “valor”, en contraste con la mera utilidad que el agente espera maximizar cuando intenta satisfacer sus preferencias meramente reveladas, a las que el filósofo canadiense también denomina “conductuales” (pp. 27-28), y que serían aquellas que sólo se relacionan con el aspecto conductual de la elección (*related to behaviour*). Así, para nuestro autor, existe una clara diferencia entre maximizar la *utilidad* de las preferencias aun sin que éstas hayan sido objeto de evaluación por parte del agente, y maximizar el *valor* de las preferencias cuando previamente se las ha sometido a examen. Por esto, y he aquí la diferencia cualitativa *i.e. normativa* que le interesa introducir a Gauthier, la conducta propiamente *racional* consiste en la maximización del *valor* de las preferencias *consideradas o actitudinales*. Este valor, pues, no es lo mismo que la mera utilidad; presenta un *plus*

³³ Sobre este “largo etcétera”, creo que resulta muy ilustrativa la crítica de Ripstein (2001) a la extensa lista de condiciones que cualificarían racionalmente una preferencia para que merezca el título de “considerada”. Esta estrategia de introducir algunas de estas condiciones, concretamente, aquellas que se refieren a la cualificación *epistemológica* de las preferencias, la encontramos no sólo en Gauthier, sino también autores como Harsanyi. A propósito de este último, Sen y Williams (1982) advierten en dicha estrategia una suerte de maniobra de “idealización” llevada a cabo, tanto por parte del utilitarismo como, en general, por parte de filosofías que –v.g. la de Gauthier, para el caso que nos ocupa– susciben el paradigma de la agencia y de la racionalidad humanas entendidas como *exclusivamente maximizadoras*.

normativo frente a ella, si bien y por mor de economía del lenguaje, Gauthier seguirá hablando en términos de “utilidad”, aunque entendiendo por ésta aquello que el agente busca maximizar cuando intenta satisfacer únicamente sus preferencias consideradas.

Ambos grupos de preferencias –actitudinales y conductuales– pueden ser consistes entre sí y, en este caso, lo que el agente *dice* preferir coincide con lo que *de hecho* prefiere en sus elecciones. Pero también puede ocurrir que ambos grupos de preferencias entren en conflicto y, entonces, lo que el individuo dice preferir no resulta ser lo mismo que termina por elegir en su conducta efectiva de elección. Si ocurre esto último, entonces puede decirse que el agente no busca maximizar el mismo tipo de utilidad cuando intenta dar satisfacción a sus preferencias actitudinales, que cuando busca satisfacer sus preferencias conductuales. Por tanto, cuando se presenta este tipo de inconsistencia se puede concluir que el sistema de valores del individuo es confuso o irracional, o que la persona no elige racionalmente –o es un agente irracional–, dado que su sistema de preferencias es inconsistente. De este modo hemos arribado, según Gauthier, a la instancia *normativa* que nos permitirá *juzgar la racionalidad* o irracionalidad de las preferencias y la conducta de elección de los agentes.

No obstante estas distinciones introducidas por el autor, en este punto creo que aún puede insistirse en preguntar: ¿qué es aquello que permite que sólo algunas preferencias sean “consideradas” y tengan, por tanto, un peso normativo frente a las demás? La respuesta de Gauthier a esta cuestión, con lo que llevamos hasta acá, aparentemente no ofrecería mayores problemas. Sin embargo al profundizar un poco más en ella y al cotejar lo ya expuesto con lo que sigue de *La moral por acuerdo*, aparecen interrogantes que en mi opinión no dejan claro este asunto de la normatividad. Nuestro filósofo va a reafirmar que la racionalidad o irracionalidad de las preferencias es independiente de su *contenido* y de los *finés* que se busque maximizar con ellas (o de los cuales ellas sean una expresión de compromiso). Para Gauthier, la racionalidad de las preferencias sólo depende de su *consistencia* lógica en tanto que *sistemas* de preferencias, así como de la cualificación epistemológica de las creencias asociadas a dichas preferencias. De allí que el autor insista en que no podamos juzgar como racional ni irracional *una* preferencia aislada, por ejemplo, aquella que se expresa en la famosa frase de Hume: “No es contrario a la razón el preferir la destrucción del mundo entero a

tener un rasguño en mi dedo”.³⁴ En este orden, según Gauthier, tampoco podríamos juzgar como irracional al individuo que exprese esta preferencia específica, por lo menos no sin antes haber examinado las relaciones de consistencia o inconsistencia que pueda haber entre dicha preferencia y el resto de preferencias que tenga el agente en cuestión.

Not the particular preferences, but the manner in which they are held, and their interrelations, are concern of reason. Once more we find ourselves in agreement with Hume, in this case when he says that it is ‘not contrary to reason to prefer the destruction of the whole world to the scratching of my finger’. It may be contrary to reason to hold such a preference in an ill considered manner, or to conjoint it with certain other preferences. But considered in it self we cannot assess its rationality. [. . .] Ends may be inferred from individual preferences; if the relationship among these preferences and the manner, in which they are held, satisfy the conditions of rational choice, then the theory accepts whatever ends they imply (pp. 25-26).

En mi opinión, estas últimas afirmaciones sorprenden, ya que resulta difícil sostener que la racionalidad o irracionalidad de las preferencias no se debe ni a su *contenido*, ni al compromiso del agente con ciertos *finés*, compromiso del que dichas preferencias son una expresión, si al mismo tiempo se sostiene, como lo ha hecho Gauthier, por una parte, que las preferencias deben ser *enunciadas* a fin de *evaluar eso que ellas enuncian*, lo cual no veo cómo puede ser dissociado de su *contenido*. Por otra parte, pienso que lo afirmado en la cita también resulta insólito si antes el autor ha sostenido que aquello que cualifica o hace racionales a las preferencias depende, en buena medida, de su base *epistemológica*, la cual tampoco veo cómo puede ser evaluada sin que al mismo tiempo se evalúe el *contenido* de las preferencias. Del mismo modo, tal vez resulte difícil defender la tesis de que la racionalidad o irracionalidad de las preferencias sólo puede determinarse luego de que se establezca la *consistencia* o *inconsistencia* que haya entre ellas, en tanto que *sistemas* de preferencias, si al tiempo que se reclaman criterios *normativos* que vayan más allá de la mera consistencia lógica con la que se contentaban, según la crítica que les hace el mismo Gauthier, los clásicos de la teoría de la decisión.³⁵ Tampoco creo que resulte fácil, por lo menos no a primera vista y sin que medie alguna explicación adicional, considerar irrelevante/ilegítima una

³⁴ *Tratado de la naturaleza humana*, libro II, parte 3, sección iii, § 416. Traducción de Félix Duque.

³⁵ Como veremos en breve, este asunto es el que parece preocupar más a autores que, como R. Brandom (2001), echan en falta el aspecto *normativo* en la propuesta de Gauthier.

evaluación de los *fin*es cuando se intenta evaluar la racionalidad de las preferencias, si al mismo tiempo se insiste, como lo hace nuestro filósofo, en que la racionalidad consiste en la maximización de un *valor*. Para aclarar un poco más este último punto, centrémonos por ahora en la discusión en torno del concepto de valor.

2.3 La noción de valor y el subjetivismo de los valores defendido por Gauthier. La tensión entre dicho subjetivismo y el interés por lo normativo

Gauthier nos previene de ser cautelosos en cuanto al uso del término “valor”, pues no ha de llevarnos al equívoco de que él crea en la existencia de valores sustantivos u objetivos –los cuales, a su entender, serían una quimera– ni, por tanto, en la existencia de fines que sean racionales o irracionales por sí mismos. Para nuestro autor, algo es un valor sencillamente porque se le busca maximizar con nuestras preferencias consideradas. Pero no ocurre al revés: la cualificación de éstas no se debe a que sean la expresión de un –supuesto– valor objetivo. En otras palabras, las preferencias consideradas son el referente gracias al cual se establece que algo es un valor; éste es deducido a partir de ellas. Por tanto, resulta ilegítimo y engañoso suponer de entrada la supuesta existencia de un valor “objetivo”, para luego concluir que éste debe ser el objeto de preferencias consideradas. Vistas así las cosas, Gauthier cae en cuenta de que tendría que responder a una pregunta ineludible: ¿cómo puede, entonces, explicarse que sea tan extendida la creencia en los valores? La respuesta del filósofo canadiense es que los valores son un producto, no un punto de partida ni un marco de referencia; se deducen de la evaluación que hacemos de las muchas formas en que *nos afectan los estados del mundo*.³⁶ Nuestro autor cree que cuando valoramos algo de manera positiva, lo hacemos porque hemos experimentado que satisface nuestros intereses y deseos, los cuales se expresan en nuestras preferencias y conductas de elección. Sin embargo, insiste también Gauthier, lo anterior no debe hacernos perder de vista que, a diferencia de la mera *utilidad*, el *valor* es aquello que se intenta maximizar mediante elecciones que obedecen a esa subclase especial de preferencias, las cualificadas o consideradas, que son las que poseen un peso normativo frente al resto, es

³⁶ El lector ya habrá advertido una clara influencia humeana, no sólo en esta respuesta puntual de Gauthier, sino también en muchos apartes de su argumentación aunque, como se verá, la posición del autor canadiense no siempre concuerda con la del filósofo escocés.

decir, frente a las meras preferencias sin evaluar. De este modo, el filósofo canadiense advierte que aunque él se adhiera a una posición netamente subjetivista en cuanto a la naturaleza de los valores, sin embargo no se debe confundir esta posición suya con un relativismo frívolo desde el cual se afirme que podemos considerar como un valor a cualquier cosa que nos produzca placer, o que sea objeto de nuestros deseos. Por tanto, tampoco se debe negar el peso normativo que tienen las preferencias consideradas con las que se busca maximizar valores.

Objects or states of affairs may be ascribed value only in so far as, directly or indirectly, they may be considered as entering into relations of preference. Value is then not an inherent characteristic of things or states of affairs, not something existing as a part of the ontological furniture of the universe in a manner quite independent of persons and their activities. Rather, value is created or determined through preferences. Values are products of our affections. [. . .] Subjectivism is not to be confused with the view that values are arbitrary . . . Values are not mere labels to be affixed randomly or capriciously to states of affairs, but rather are registers of our fully considered attitudes to those states of affairs given our beliefs about them. *Although the relation between belief and attitude is not itself open to rational assessment* is not therefore arbitrary (pp. 46-48; las cursivas son mías).

En esta última cita me he permitido subrayar la tesis de que no sería *racionalmente evaluable* la *relación* entre las *creencias* que proporcionan la base epistemológica de nuestras preferencias y aquellas *actitudes* nuestras que se expresan en nuestros *juicios de valor*, pues creo que no resulta fácil hacer compatible esta tesis con la importancia que el mismo Gauthier dice asignarle a la *cualificación epistemológica* de las mencionadas creencias. Igualmente, a mi entender, dicha cualificación tampoco parece ser compatible, ni con la poca importancia que ahora el autor le otorga al *contenido* de las preferencias, ni con su prohibición de asignarle algún sentido *normativo* a los valores. Por todo ello no creo que nuestro filósofo responda a la siguiente objeción: si tal y como él mismo lo había sostenido, resultaba tan importante la *cualificación epistemológica* de las creencias asociadas a las preferencias, entonces resultaría muy difícil que ahora se afirme, sin más, como lo hace el autor, que el *contenido* de las preferencias *no es evaluable racionalmente*, sino que sólo pueden serlo los *nexos formales* entre las preferencias. Pienso que, por ende, sería insólito que ahora Gauthier advierta que no puede llevarse a cabo un *examen racional* de la *relación entre*

creencias y actitudes, relación que tendría que ser valorada como fundamental para que podamos tener preferencias *consideradas*.

Creo que las dificultades que aquí señalo se deben, en último término, a un doble intento que, de no ser visto como estrictamente contradictorio, al menos sí sería difícil de sostener: 1) De un lado, nuestro autor busca establecer unos estándares *normativos* para las preferencias consideradas, estándares exigentes desde el punto de vista epistemológico y que irían más allá, repito, de la mera consistencia lógica con la que se contentaban, según el mismo Gauthier, los teóricos de la decisión. 2) Pero de otro lado y en clara tensión con ello, están las reticencias del filósofo canadiense a asociar dichos estándares con los *contenidos* de tales preferencias, o con los *fin*es que el agente busque al formularlas. En otros términos, por una parte Gauthier afirma que debe haber referentes *normativos* que permitan *evaluar* la *racionalidad de las preferencias* y de los agentes (o de su conducta de elección). Pero, por otra parte, insiste en rechazar referentes normativos que *no sean reducibles a meras preferencias* o que, finalmente, no colapsen en meras preferencias. Para satisfacer ambos propósitos suyos, el autor intenta asignarle a las preferencias mismas un peso normativo, mediante el establecimiento de criterios que las cualifiquen. Con esta maniobra se lograría que por lo menos algunas de ellas admitan ser vistas como el producto de una reflexión y de un compromiso cualificado por parte del agente; compromiso que se hace posible gracias a las circunstancias en las que éste se las ha formulado.³⁷ Esto implica que el individuo tendría que ser conciente de tales preferencias, para lo cual es necesario que las haya verbalizando y que para hacerlas suyas haya contado con la información relevante, así como con las ya mencionadas condiciones para la formulación de creencias plausibles que puedan dotar de una buena base epistemológica a las creencias del agente. De esta manera, los estándares normativos con los que deben cumplir las así llamadas preferencias consideradas, para que cuenten como tales, se resumen en que ellas sean precisamente eso: “consideradas”. Con esto, como puede ver el lector, o bien se concluye que estamos ante una *petitio principii*, o bien habría que admitir que no son

³⁷ Como de algún modo se anunció en un pie de página anterior, Harsanyi (1982 y 1985) se ve enfrentado al mismo problema y su estrategia de solución es bastante parecida a la que aquí propone Gauthier: se busca dotar de un peso normativo a un subgrupo especial de preferencias, pero se evita identificar dicho subgrupo con referentes *normativos* que *no* pertenezcan a la clase genérica de las *preferencias*. Esto es, se evita hablar de *normas y valores* a los que se apele cuando se evalúen las preferencias, y cuya naturaleza sea distinta o signifique un *plus* frente a lo que podríamos llamar la “clase natural” de las preferencias.

claros los criterios que propone Gauthier para calificar de “consideradas” a esta exclusiva subclase de preferencias.

Al llegar aquí podríamos preguntarnos: ¿en virtud de qué el agente asume conciente y reflexivamente dichas preferencias? De nuevo, la respuesta de Gauthier prohíbe que se apele al *contenido* de tales preferencias, o a los *fin*es que busca el agente. Lo cual se hace aun más confuso, ya que nuestro filósofo insiste en hablar de *valor* en vez de mera *utilidad*, al tiempo que recalca la diferencia cualitativa que separa a las preferencias *verbalizadas* de aquellas otras que simplemente son *reveladas* por la mera *conducta* de elección. Gauthier intenta ser tan radicalmente fiel a su consigna de no concederle importancia a los *contenidos* expresados en esta verbalización de las preferencias, que creo que con ello en ocasiones su intención normativa se hace cada vez más difícil de sostener. Volvamos de nuevo a la famosa frase de Hume. El filósofo canadiense afirma que, frente a la persona que sostenga que prefiere la destrucción del mundo a sufrir un rasguño, podríamos válidamente lanzar el juicio de que no está en sus cabales y que, por tanto, acaso sufra una suerte de desorden orgánico, incluso equiparable a una indigestión (p. 49). Pero de ningún modo estaríamos autorizados a afirmar que se trata de un agente irracional ni, mucho menos, podríamos para ello invocar nuestros valores, dado que éstos (como todos los valores) son meramente subjetivos, pues expresan nuestros sentimientos o afecciones frente a aquello que puede ser objeto de nuestras preferencias. La evaluación de estas últimas debe remitir a un examen de su base epistemológica, pero este examen se refiere al conocimiento que el agente tiene de los hechos relevantes para la formulación de tales preferencias; conocimiento que según Gauthier, es necesariamente empírico: no se trata en absoluto de una suerte especial de conocimiento no-empírico o intuitivo cuyos objeto fuesen los valores o los fines perseguidos por los agentes.

Como ya se ha dicho, para nuestro autor los fines y los valores no poseen un estatus de objetividad en virtud del cual ellos pudiesen proporcionar los estándares normativos que permitiesen juzgar o evaluar las preferencias de los agentes. Antes, por el contrario, al ser deducidos de éstas, los valores son objeto del mismo tipo de conocimiento que empleamos para evaluarlas a ellas. De esta manera según Gauthier, se puede asegurar el peso normativo de las preferencias consideradas y con ello una posición no-relativista o no-irracionalista frente a éstas, al tiempo que se puede y se

debe suscribir el viejo *dictum* –empirista y subjetivista– humeano de que no puede juzgarse de irracional ninguna preferencia, incluyendo v.g. la de “preferir la destrucción del mundo a sufrir un rasguño”. Pues el contenido de esta preferencia específica o el fin que persiga el agente que la suscriba no son, en sí mismos, susceptibles de examen racional. A lo sumo, repito, y sólo después de que se hayan considerado otros hechos relevantes en relación con el agente v.g. su estado de salud, podemos afirmar según Gauthier, que dicho agente no está en su sano juicio. Sin embargo, un dictamen como éste no puede basarse en el contenido de su –tal vez excéntrica– preferencia, ni debe referirse a la racionalidad o irracionalidad de ésta, sino sólo a un desorden en las afecciones del sujeto. Empero, estas últimas, según nuestro autor, al ser independientes de la capacidad racional del agente, *no* son susceptibles de examen racional alguno; del mismo modo en que tampoco se examinan racionalmente los desórdenes digestivos.³⁸

Although . . . we might have grounds independent of the content of this preference, for holding such a person to be mad, we should not have grounds for considering his preference arbitrary. Madness need not imply either a failure of reason or a failure of reflection. (Madness we should hold, is primary a disorder of the affections. But this does not imply that the affections are irrational, any more than a disorder of the stomach implies that it is irrational.) [. . .] Subjectivism is also not to be confused with the view that values are unknowable. Evaluation, as the activity of measurement, is cognitive. Preference, what is measured, is knowable. What the subjectivist denies is that there is knowledge of value that is not ordinary empirical knowledge, knowledge of a special realm of the valuable, apprehended through some form of intuition differing from sense-experience. Knowledge of value concerns only the realm of the affects; evaluation is cognitive but there is no unique ‘value-oriented’ cognition (pp. 48-49).

No obstante estas últimas afirmaciones suyas, creo que sigue siendo difícil que Gauthier asegure un estatus normativo para las preferencias consideradas. Nuestro autor trata de salvar este estatus, pero en su intento por separarse de un supuesto “objetivismo” de los valores,³⁹ reduce todo valor a mera expresión de preferencias, con

³⁸ El lector habrá advertido que en esta afirmación se supone una tesis controvertible e importante (pero de la que por ahora no puedo ocuparme) sobre la imposibilidad de un contenido cognitivo asociado a aquello que Gauthier llama “afecciones”. Véase al respecto: Hampton, Jean (1988), Murphy, Jeffrie (1988); Rorty, Amèlie (1980a); Stocker, Michael (1980); De Sousa, Ronald (1980); Greenspan, Patricia S. (1980).

³⁹ El único autor “objetivista” que menciona Gauthier es G. E. Moore (*Cf.* trad. 1997). Curiosamente, no menciona a Scheler (*Cf.* Trad. 1941). Creo que el lector convendrá en que no resulta muy acertada esta preocupación por el objetivismo, dado que propuestas como la de Moore o la de Scheler no parecen ser suscritas en serio por las figuras más representativas de la filosofía moral contemporánea, especialmente en lo que se refiere al sentido en que tendría que entenderse el término “valor”. Creo que, de hecho, una

lo cual, como hemos visto, caeríamos o en una argumentación circular o en una contradicción. En este punto, pienso que habría que atender a la objeción que plantean Brandom (2001) y Ripstein (2001): sólo podemos evitar caer en esta argumentación circular si se piensa a los estándares *normativos* como algo que *no* puede ser reducible a *meras preferencias*. Se los debe concebir, entonces, como aquello que explica, precisamente, el que algunas de estas preferencias merezcan el título de “consideradas”. Esto es, como aquello que permitiría que hubiese una *evaluación* de las preferencias. Más exactamente, los estándares normativos serían esa instancia a la que se apela cuando se hace tal evaluación. Pero Gauthier parece estar más preocupado por librar una batalla contra lo que él llama posiciones “objetivistas” frente al tema de los valores o contra un “intuicionismo” de los valores, que por contestar este tipo de objeciones. Y en esta batalla contra molinos de viento, termina por atacar al cognitivismo ético que creo que tendría que suscribir de alguna forma, si no quiere caer en posiciones emotivistas como aquellas de las que él querría separarse.⁴⁰

Si las “afecciones”, tal como las llama nuestro autor, no tienen contenido cognitivo de ningún tipo⁴¹ —hasta el punto de compartir una naturaleza común con los desórdenes digestivos—, y si los valores son expresión de tales afecciones, entonces éstos pertenecen al reino de lo irracional; reino que, en mi opinión, en el discurso de Gauthier parece hacerse cada vez más amplio frente al cada vez más estrecho reino de una racionalidad que, por lo visto, quedaría reducida a muy poco, tal vez a mera consistencia lógica. Por supuesto, nuestro autor no estaría en absoluto de acuerdo con

posición “objetivista” resultaría ser peregrina si se considera el contexto en el que se plantean en nuestros días las discusiones en filosofía moral, sobre todo después del giro lingüístico. Si algunos autores contemporáneos mencionan el término “valor”, no lo hacen suscribiendo una posición objetivista, ya tan extraña, repito, para un lector actual. Por ejemplo, y como se verá en el capítulo quinto, Ch. Taylor, habla de “compromisos valorativos”, pero no le asigna un estatus de objetividad a aquello que dichos compromisos expresen o defiendan, sino que sitúa la articulación de tales compromisos en el contexto del lenguaje y de lo simbólico; de lo que las cosas *significan para* los agentes y del carácter *dialogico* de los procesos en los cuales *se construyen los significados compartidos* y se narran o expresan las experiencias con-vividas por dichos agentes. Así mismo, A. Sen (1992) se refiere a los “compromisos” (*commitments*) que tiene un agente, compromisos que, según el economista indio, explican el hecho de que, en no pocas ocasiones, un sujeto, sin dejar de ser racional, vaya más allá e incluso actúe en contra de algunos de sus deseos y preferencias.

⁴⁰ Gauthier (pp. 2, y 55 y ss.) hace una crítica al emotivismo que le atribuye a pensadores como Mackie, Stevenson y, como veremos, también a Parfit (pp. 38, 174, 184, y 342-343).

⁴¹ Como arriba se ha mencionado en un pie de página, si bien no puedo detenerme ahora en este problema del contenido cognitivo de las afecciones, creo que resulta importante plantearlo cuando se intenta identificar los supuestos que estarían a la base de cualquier modelo de agencia humana.

estas críticas. Muy al contrario, intenta mostrar la amplitud y complejidad de los terrenos de la racionalidad tal y como él la concibe. Veamos entonces, a continuación, cómo aparecen los contornos de dicha racionalidad en la propuesta de Gauthier y, en estrecha relación con ello, de qué manera, para apuntalar mejor su tesis de que la moral se justifica a partir de razones utilidad, nuestro autor propone al *mercado* como modelo de la interacción moral-racional y, por tanto, a la agencia económica como modelo de agencia humana.

3. La “arena de la moral”: cooperación, autointerés y agencia económica

Con lo dicho hasta acá parecería que el modelo de racionalidad práctica asumido por Gauthier se limitaría al de una racionalidad paramétrica ejercida por un agente que actúa en solitario. No obstante nuestro autor advierte que él está pensando en una racionalidad estratégica, es decir, aquella que orienta el tipo de decisiones en las que debe tenerse en cuenta cómo las propias elecciones y expectativas son afectadas por (y afectan a) las elecciones y expectativas de los demás agentes. Esta es, pues, para Gauthier, lo que él llama la “arena de la moral” (pp. 60 y ss). Se trata del ámbito de interacción que nos es propio en tanto que agentes racionales vinculados mediante ciertos constreñimientos morales que, a su vez, serían el instrumento con el cual se intentaría restringir la tendencia de cada individuo a maximizar su propio beneficio, restricción que tiene como finalidad que se alcance una situación estable o deseable para todos los sujetos así relacionados. Nuestro autor afirma que aunque se hable aquí de una interacción entre maximizadores, el hecho de que dicha interacción también esté mediada por constreñimientos morales hace que se la pueda pensar, así mismo, como un juego cooperativo en el que se busca armonizar dos tendencias racionales que no siempre son conciliables: la tendencia al óptimo y la tendencia al equilibrio⁴²: “Moral theory is essentially the theory of optimizing constraints on utility-maximization”. (p. 78). Según el filósofo canadiense, la armonía que imponen las restricciones morales a las interacciones humanas se produce de tal manera que estas últimas se rigen por una

⁴² Sobre las nociones de “óptimo” y “equilibrio”, véase: Resnik (1998, pp. 203-261).

estructura esencialmente *cooperativa*, lo cual arroja el resultado opuesto al que se obtiene de una situación de dilema del prisionero (p. 79).

Recordemos que, como ha sido comentado por tantos autores, en una situación de esta naturaleza cada agente no tiene, al parecer, otra alternativa distinta a la de buscar su propio beneficio *i.e.* elegir “racionalmente” en el sentido de maximizar su utilidad esperada; con ello, e incluso sin que sea conciente de ello, el sujeto estaría decidiendo de acuerdo con los cánones de la teoría de la elección racional. Sin embargo, lo interesante de un dilema del prisionero es que, a pesar de la racionalidad desplegada por cada agente y, por ende, aunque cada cual opte por aquella elección que considera la “mejor” o “más racional” (siguiendo los mencionados cánones), dicha racionalidad es el factor que paradójicamente termina por producir una situación final que resulta ser *indeseable* para todos los implicados. Esto es, un resultado que estaría lejos de considerarse como “elegible” o “racional” de suyo.⁴³ En otros términos, lo característico de esta indeseable situación resultante es que se obtiene a causa de una estructura de interacción *no-cooperativa* a la que parecen “estar condenados” los agentes que allí participan, no pudiendo escapar de ella justamente en virtud de su ser-*rationales*; con lo cual, repito, podría parecer que estamos ante una suerte de aporía.

Frente a este conocido quebradero de cabeza, Gauthier cree que uno de los mayores aciertos de su teoría está precisamente en que permite hacer plausible, tanto un esquema de interacción, como un resultado completamente opuestos a los que se ven avocados los agentes involucrados en una situación de dilema del prisionero. Nuestro autor cree que él sí logra mostrar la plausibilidad de unas circunstancias de interacción que, al contrario de lo que sucede en la indeseable situación del dilema, llevarían a cada agente a adoptar una actitud *cooperativa*, dado que el sujeto prevé que ello le significa un precio *razonable* –renunciar a parte de su utilidad esperada–; precio que está dispuesto a pagar en vista de los *beneficios* que espera obtener al participar en el esquema cooperativo de interacción, tal y como éste se propone en *La moral por acuerdo*. Partiendo de dicho esquema, cada uno de los agentes así relacionados *motu proprio* se movería a colaborar para el logro del beneficio de todos. Cada uno estará dispuesto, sin que a ello le empuje coerción alguna, a restringir en parte su conducta

⁴³ Sobre este tan conocido asunto del dilema del prisionero, véase, por ejemplo: R. Nozick (1995); R. Axelrod (1984), y Ph. Pettit (1986).

maximizadora, ya que sabe que al mismo tiempo lograría una ganancia personal a la que sólo puede acceder en tanto se obtenga el beneficio de todos los involucrados.

De esta manera, según el filósofo canadiense, la moral reemplazaría a la mano invisible a la que se refiere A. Smith para tratar de explicar la armonía social como producto final de las interacciones entre individuos que, al buscar cada uno su propio beneficio (o el de su familia), contribuyen sin proponérselo a la producción del bien general.⁴⁴ Si para Gauthier la moral es lo que explica tal armonía, entonces ella es lo que permite que el resultado de la interacción entre agentes racionales, a diferencia del resultado de una situación de dilema del prisionero, sea “racional” y en él coincidan óptimo y equilibrio. Nuestro autor afirma que esta deseable situación también puede esperarse como el producto del funcionamiento de un mercado que esté libre de aquellas imperfecciones a causa de las cuales unos agentes podrían ser explotados por otros, o se darían transacciones forzosas en contra de los intereses de quienes lleven la peor parte. Veamos, pues, cómo establece Gauthier este vínculo entre mercado y moralidad.

3.1 El mercado, sus agentes egoístas y por qué la moral reemplaza la mano invisible. Las tensiones entre el mercado sin fallos y los mercados reales

Si bien Gauthier reconoce el poder metafórico de la mano invisible, no obstante también advierte que habría una diferencia fundamental entre la situación que (mediante dicha imagen) describe Smith, y lo que sucede con el sistema de moralidad que se propone en *La moral por acuerdo*. En el caso de un mercado perfecto que pareciera estar guiado por la mano invisible, no habría razones para que los agentes acepten restricciones morales a la búsqueda de su propio interés, restricciones que, recordemos, son las que intenta justificar el filósofo canadiense mediante su estrategia de mostrar que éstas nacen del autointerés de los agentes racionales y no concernidos por sus congéneres. De allí que nuestro autor considere que un mercado sin fallos debe ser pensado como un terreno moralmente neutro *i.e.* una zona en la que la moral no jugaría papel alguno, puesto que quienes allí participan no serían víctimas de ninguna forma de explotación; por lo que en ese tipo de escenario no tendría sentido intentar poner freno a la actitud maximizadora de los agentes involucrados. Empero, se pregunta Gauthier,

⁴⁴ Cf. Smith (Trad. 1982).

¿qué sucede si tal mercado no existe, y si lo que hay en realidad son mercados que presentan imperfecciones que harían deseables los constreñimientos morales, a fin de prevenir las injusticias y perjuicios originados en tales imperfecciones?

La respuesta del canadiense viene dada por una tesis que resulta fundamental dentro de su texto: las restricciones morales a la búsqueda del propio interés se hacen necesarias porque, en ausencia de éstas, se presentarían precisamente esas situaciones de explotación y desventaja debidas a los fallos del mercado: “Morality arises from market failure” (p. 84). La moral, pues, se hace indispensable, según el autor, allí donde haya imperfecciones del mercado, e innecesaria e injustificada en aquellos mercados que sean perfectamente competitivos.⁴⁵ Por lo anterior, Gauthier expresamente anuncia que su estrategia para fundamentar su propuesta de una moral por acuerdo consistirá, en primera instancia, en describir cómo habría que pensar esa zona moralmente neutra en la que no estarían justificadas las restricciones morales; es decir, un sistema de interacciones que funcione como un mercado perfecto. Y luego, en segundo término, el filósofo canadiense mostrará por qué se haría necesaria la moral en un escenario totalmente opuesto, es decir, dentro de aquel que presente fallos. La tesis de nuestro autor, así como esta estructura argumentativa con la que promete justificarla, en mi opinión, amén de ser plausibles y bastante trajinadas, no tendrían en principio por qué despertar reticencias.⁴⁶

Sin embargo creo que dicha estrategia argumentativa no es desplegada de manera clara ni en el orden anunciado por Gauthier, amén de que éste introduce elementos que pueden hacerla aparecer como inconsistente. En muchos apartes de *La*

⁴⁵ El lector habrá advertido que si esto es así, entonces pueden caer sombras de duda sobre la compatibilidad de esta tesis con otra afirmación importante que hiciera Gauthier al inicio de su texto, y en la que nuestro autor resume su propuesta de fundamentación de la moral. A saber, que las normas morales —de una moral propiamente filosófica, no tradicional, o sea, racionalmente justificada— son un subconjunto de una clase genérica más amplia conformada por las reglas de elección racional. Estas últimas son las que guiarían las decisiones de agentes maximizadores, a los que nuestro autor cree que debe pensarse como autointeresados y no concernidos por el bienestar de sus congéneres, exactamente como se piensa a los agentes económicos que transan en un *mercado*. Con lo cual, como puede verse, habría aquí una afirmación que no parece conciliarse fácilmente con la nueva tesis que ahora busca demostrar Gauthier: que la moral se hace necesaria como la *respuesta racional a las imperfecciones del mercado*. Pues, en primer lugar, se trata del sistema conformado precisamente por ese mismo tipo de agentes y, en segundo término, se hace difícil establecer de qué forma de “racionalidad” estamos hablando.

⁴⁶ De hecho es una de las tesis fundamentales, tanto del contractualismo en general, como del tipo de liberalismo de corte más “igualitario”, el cual, paradójicamente, Gauthier descalifica en muchos apartes de su obra, sobre todo cuando critica a autores como Rawls, que se oponen a un liberalismo de carácter más “libertario”, tal y como aquel con el que simpatiza Gauthier.

moral por acuerdo se advierte, por un lado, lo que acaba de afirmarse con respecto al mercado perfecto o ideal: que no existe y que por ello se hace necesaria la moral, con el fin de prever o solucionar los problemas causados por los fallos de los mercados reales, que son los que verdaderamente existen. Empero, en segundo lugar, el canadiense así mismo se empeña en postular como un *ejemplo a seguir* el modelo de una sociedad que funcione a la manera de un *mercado perfecto* guiado por la mano invisible, esto es, una sociedad en la cual sería *innecesaria la moral*, o en la que no se haya diseñado ex profeso un sistema de incentivos para seguirla o, mejor aun, donde estos últimos aparecerían como no justificados a los ojos de quienes participen en un mercado de tal naturaleza, ya que en él nadie es víctima de explotación. Si se coteja el mencionado anuncio que hace Gauthier de su estrategia argumentativa con lo que sigue de su exposición, creo que en ella se introduce un elemento que la vuelve un tanto confusa: el *elogio* de esta zona moralmente neutra, pero no tomada como un mero ideal, sino como el ejemplo a seguir; como el modelo que *debería* aplicarse en toda sociedad real o posible.

El problema es que este elogio se funda en razones morales. Nuestro autor considera al mercado perfectamente competitivo como *el mejor* esquema de interacción desde el punto de vista *moral*, lo cual confunde aún más, dado que anteriormente se ha reiterado que en dicho esquema de interacciones *no* tiene cabida la moral, o no serían deseables las restricciones morales a la conducta maximizadora, en vista de que tales restricciones no estarían justificadas –i.e. serían irracionales– tratándose de un mercado en el que este tipo de conducta no produce situaciones de explotación.⁴⁷ Si se busca demostrar la necesidad de una moral para maximizadores egoístas, así como la razonabilidad que éstos le atribuirían, razonabilidad que explica el que estos agentes voluntariamente se comporten de acuerdo con las normas que dicha moral les imponga, entonces creo que en este punto se le podría preguntar a Gauthier: ¿qué es aquello que puede, pues, modelar el modelo de un mercado ideal que no requiere de restricciones morales, o que es moralmente neutro? Amén de lo anterior, ¿qué relevancia *moral* tendría dicho modelo, sabiendo que tal mercado ideal se identifica, o bien con una zona

⁴⁷ En este sentido pienso que a lo largo de su discurso Gauthier se mueve entre las dos formas de defensa moral del mercado a las que se refiere Félix Ovejero: 1) En ocasiones señala las buenas *consecuencias* que trae el ejercicio de la libertad económica y, 2) Otras veces apela al *valor* de dicha *libertad*, valor que pone por encima de consideraciones sobre las consecuencias que dicha libertad pueda traer. Cf. Ovejero (1994a y 1994b).

pensada como *no* necesitada de moral, o bien con una situación irreal que es claramente contradicha por mercados que sí existen y que acusan situaciones moralmente indeseables?

Aun suponiendo que, precisamente por esto último, se entienda la pertinencia metafórica y pedagógica de apelar a dicho modelo, ¿cómo se puede desde allí lograr que se haga plausible y/o útil la pintura opuesta, esto es, la de una zona moralmente regulada? Si nuestro autor se propone justificar las restricciones morales dentro de un contexto en el que interactúan maximizadores egoístas, pienso que tal vez sería más aconsejable comenzar reconociendo los límites morales de dichas interacciones y de la actitud de este tipo de agentes no-concernidos por sus congéneres. Por ende, el punto de partida no tendría que ser el elogio moral sino la crítica a los mercados reales y al *ethos* egoísta que éstos puedan propiciar. Así, una vez reconocidos los problemas presentes en dichos mercados, podría luego mostrarse por qué en ellos se hacen necesarias/deseables las normas morales, a fin de evitar los daños producidos por las imperfecciones del mercado, imperfecciones agravadas por la actitud de los agentes maximizadores egoístas. Sin embargo, no es ésta la estrategia seguida por el filósofo canadiense, quien, repito, comienza por elogiar *moralmente* una zona *no* necesitada de moral, proponiéndola como el *ejemplo* a seguir por parte de toda sociedad humana, para terminar con un movimiento argumentativo bastante curioso y ante el que cabe la sospecha de un interés ideológico por parte de Gauthier: mostrar al mercado como institución moderna, y a la moderna sociedad de mercado –en tanto que eventos *reales/históricos*, y *no* en tanto que meros *modelos conceptuales*–, como los ejemplos de *perfección moral* que *deberían* seguir todas las sociedades actuales o posibles (pp. 99-101).

Otro de los elementos que puede contribuir a esta confusión a la que aludo, tal vez se deba al hecho de que la referencia que hace el canadiense a la mano invisible de Smith, la convierte en una metáfora difícil de manejar en *La moral por acuerdo*. Ya es un lugar común que el clásico escocés acude a esta figura para explicar la armonía artificial de intereses que, según él, parece producirse a pesar de que los individuos que detentan dichos intereses persigan su propio bienestar o el de sus familias, y no contribuyan de una manera conciente o intencionada al logro de un bienestar general de la sociedad. Sin embargo, también creo que no debería perderse de vista, como en

cambio parece que en ocasiones se desdibuja en Gauthier, que en Smith la mencionada armonía de intereses no puede surgir sin que previamente se hayan establecido unas reglas de juego para el buen funcionamiento, no sólo del mercado, sino de las interacciones sociales en general. El cumplimiento de tales reglas es lo que permite que surja y se sostenga el mercado, de modo que las primeras no son el producto de dicho sistema sino que, por el contrario, son aquello (el marco restrictivo) que lo hacen posible y condicionan al mercado como un terreno de cooperación no-forzosa y menos riesgosa.

Creo que en este punto, y como se irá viendo en lo que sigue, un lector puntilloso podría objetarle a Gauthier el hecho de que restrinja el alcance que Smith quería darle a su idea de la armonía (no intencionada) de intereses, la cual el filósofo escocés, repito, tal vez planteaba como válida a nivel de toda la gama de interacciones sociales, mientras que el autor canadiense pareciera reducir dicha armonía al mero ámbito de las transacciones económicas. Pienso que para Smith y en contraste con el uso que hace Gauthier de la metáfora de la mano invisible, la naturaleza que puede uno atribuirles a las mencionadas reglas de interacción social no es de suyo económica, como parece entenderlo el autor de *La moral por acuerdo*. Y aun cuando se conviniera en no calificarlas como “morales”, por lo menos sí creo que se las podría considerar o bien “políticas”,⁴⁸ o bien “institucionales”, en tanto que condición de un orden económico no coercitivo.⁴⁹ Por tanto, en contraste con la pintura que muestra Smith, pienso que en Gauthier las cosas se tornan un tanto confusas, pues a veces parece que sostiene que la moral *sigue al* mercado, pero en otras ocasiones parece afirmar que ella hace parte del marco normativo que *permite*, como en Smith, que haya mercado. De

⁴⁸ Como se verá en el tercer capítulo, también dedicado a Gauthier, este es uno de los puntos en los que nuestro autor piensa que su propuesta aventaja a las teorías de Hobbes y de Rawls, a quienes el filósofo canadiense critica por ser deficitarias en cuanto a su contenido *moral* y por cargar demasiado las tintas en el aspecto *político*.

⁴⁹ En cuanto a la interpretación que Gauthier hace de Smith, insisto en que en el primero se acusan confusiones que acaso no se vean en el segundo, entre otras razones, porque considero que este último no estaría señalando al mercado como el único campo en el que se despliega la armonía social, sino más bien a dicha armonía como algo que caracteriza las interacciones sociales en general y, dentro de ellas, al mercado como un logro de lo social. No al revés: no se llegaría a lo social como un subproducto de una lógica del mercado que sea “superior” ni previo a lo social, ni que se imponga con unas reglas de juego que no sean sociales de suyo. A pesar de las muchas críticas que ha recibido (v.g. que la mano invisible es un posible objeto de fe que iría en contravía de un análisis crítico del mercado), no en vano Smith habla de “sentimientos morales” y en su discurso se advierte la influencia de sus maestros y contertulios escoceses, en una época en la que la economía aún no se había escindido de las ciencias “morales”. (Cf. Smith, Trad. 1982, y Trad. 2004); Young, Jeffrey (1997). Mathiot, Jean (1990).

todas maneras, al final lo que queda claro en Gauthier es que, si ha de haber sociedad y cooperación humanas, éstas *deben* seguir el modelo –normativo– del mercado.

Esta última afirmación hace que la disyuntiva obvia ante la cual se encuentra el lector es si debe entender que aquí el filósofo canadiense se refiere al modelo que ofrece el mercado *perfecto o ideal*, o si más bien postula como “ideales” o ejemplos a seguir a los mercados *reales*. Si se elige la primera opción, entonces se presenta el ya mencionado problema de que el mercado ideal no parece ser un hecho histórico, sino que es sólo eso: un ideal, y un ideal dentro de cuya concepción, paradójicamente, *no* entra la *moral*; con lo cual no se vería su pertinencia como modelo de sociedad justa o de cooperación humana *justa*. Y si se opta por la segunda alternativa, entonces habría que entender que aquí nuestro autor propone que se siga el “modelo” (si es que puede hablarse así de ellos) ofrecido por los mercados reales. Pero esto último nos pondría ante el inconveniente de que dichos mercados, los que sí existen, presentan fallos que propician el que unos agentes sean explotados por otros, quedando por tanto en entredicho el carácter de modelo “ejemplar” y moral que parece atribuirles Gauthier. En mi opinión, la solución a esta disyuntiva no aparece claramente expuesta en *La moral por acuerdo*. Por una parte, repito, el filósofo canadiense afirma lo que todos sabemos: que el mercado perfecto no existe. De allí la tesis que tanto le interesa demostrar: que la moral se hace necesaria dadas las imperfecciones del mercado y que, por ende, ella es la respuesta racional a dichas imperfecciones.

Creo que, insisto, esta tesis no ofrecería mayores problemas e, incluso, podríamos darle la razón a Gauthier, en vista de que los mercados reales, tal y como lo hemos dicho, no son propiamente un ejemplo de moralidad y, por ende, dicha moralidad sería la solución a la que estaría racionalmente dispuesto a contribuir cualquier grupo de agentes, incluso los maximizadores egoístas.⁵⁰ Empero, por otra parte, y de manera sorprendente, nuestro autor, repito, elogia dichos mercados reales-históricos como un auténtico ejemplo de superioridad moral, elogio que se aprecia claramente en sus reiteradas y entusiastas apologías de la sociedad de mercado que es real-histórica, y no meramente ideal. Finalmente creo que al lector no puede menos que asaltarle la duda de

⁵⁰ Haciendo, por supuesto, ciertas salvedades, ante las posibles sospechas, arriba ya mencionadas, que puedan presentarse frente a la motivación moral que pudiera tener un maximizador MI que decida convertirse en un MR; especialmente si se encuentra en circunstancias en las que puede abusar (sin sufrir retaliaciones) de su posición dominante en el mercado. Sobre este asunto se volverá más adelante.

si era realmente necesario acudir al modelo del *mercado* –bien sea ideal, bien sea fáctico–, tanto para justificar la *moral* postulándola como la clave de toda cooperación humana no forzosa, como para defender un modelo de agente en tanto que participe de dicha estructura de cooperación. Mi intuición es que Gauthier utiliza todas estas complicadas maniobras en su argumentación, como una manera de apuntalar *su modelo agente* entendido como un ser *no-concernido* por sus congéneres; como un individuo egoísta que acude a la moral únicamente en razón de su autointerés. En lo que sigue, el lector podrá juzgar si esta intuición es plausible. De manera que, por el momento, no insistiré más en estas críticas a la pertinencia de la estrategia argumentativa de nuestro filósofo y trataré de seguir su esquema expositivo, partiendo de su descripción del mercado perfecto en tanto que zona no-moral o no necesitada de moral para, posteriormente, detenerme en lo que más me interesa analizar: el tipo de *agente* que opera en ese contexto especial que proporciona el mercado.

3.2 El modelo de Robinson Crusoe. El agente económico: los problemas para su autonomía y moralidad

Al mercado perfectamente competitivo, dice Gauthier siguiendo los manuales clásicos, se lo debe concebir como careciendo de fallos. Esto quiere decir que habría que pensarlo como libre de externalidades; sin los problemas que plantea la existencia de bienes públicos; con unos derechos de propiedad claramente definidos; sin trabas a la libre actividad económica y, por tanto, sin un centro que controle dicha actividad. Se trataría de un mercado en el que ninguno de los agentes involucrados se vería forzado a relacionarse económicamente con otro(s) en contra de sus propios intereses. Allí nadie sería explotado, ya que no sería objeto de ninguna forma de traslado unilateral de costes; no habría, pues, rentistas, parásitos, ni polizones. Igualmente, no se presentarían ni monopolios ni competencia desleal; la libertad de actividad económica y la no-coacción estarían garantizadas para todos por igual, de manera que nadie saldría perjudicado por la libertad ejercida por otros agentes, garantizándose así una competencia limpia, mediante unas reglas de juego que serían las mismas para todos. Si los anteriores requerimientos se cumplen, entonces la producción y el intercambio se darían en condiciones de certidumbre y seguridad, al tiempo que coincidirían la optimalidad y el equilibrio. El resultado de esta hipotética situación es que, como arriba

se mencionó, parecería estar guiada por la mano invisible de Smith, pues cada agente, buscando su propio beneficio, contribuiría sin quererlo al beneficio de todos.⁵¹ Este panorama constituye aquello a lo que Gauthier se ha referido como una zona “moralmente neutra”, a la que cabría pensar como el *locus* en el que no serían necesarios los constreñimientos morales, en tanto que éstos se tomen como limitaciones a la búsqueda del interés individual o como intervenciones extraeconómicas que impidan el despliegue de la libre actividad económica. Dichas limitaciones tampoco estarían justificadas a los ojos de los propios agentes, ya que ellos, ante el hecho de que las reglas de juego favorecen a todos –y no sólo a algunos–, no verían razones para restringirse en su actividad maximizadora. Cada uno podría, pues, desarrollar sin trabas su libre actividad (*free activity*) con la única restricción que suponen unas leyes justas, defendidas por un aparato judicial y unas instituciones eficientes e imparciales (que hagan cumplir los contratos, e impidan el fraude y las transacciones forzosas), así como aquellas normas orientadas a salvaguardar una libre y limpia competencia económica.⁵²

Para ilustrar la importancia de la *libertad* de la que gozaría, según nuestro autor, un sujeto que se encontrara en la situación de los agentes económicos que transan en un mercado perfectamente competitivo, Gauthier acude a una figura que, en mi opinión, resulta muy reveladora del modelo de agente por el cual apuesta el filósofo canadiense: Robinson Crusoe. Si Robinson es el único habitante humano de su isla y no se encuentra con ninguna restricción a sus actividades –salvo aquellas que le imponga la naturaleza–, los beneficios que él obtenga de tales actividades dependen únicamente de sus esfuerzos y talentos, los cuales puede invertir en lo que a él le parezca. De este modo no podría culpar a nadie, más que a sí mismo, si ocurriera que no lograra obtener los beneficios que esperaba. Gauthier subraya que es esto lo que justamente le

⁵¹ Esto no quiere decir que en el largo plazo el funcionamiento de este mercado perfectamente competitivo arroje como consecuencia una igualdad de riqueza entre quienes participen en él. Garantizar la ausencia de fallos y la competencia limpia no significa garantizar dicha igualdad, ni que todos tengan el mismo éxito ni los mismos logros: seguramente a unos le irá mejor que a otros. La única igualdad que debe garantizarse es la igualdad ante la ley: un marco normativo que no beneficie ni perjudique a nadie en particular.

⁵² Posteriormente se comentará una de las imperfecciones del mercado que puede significar una desventaja de unos agentes frente a otros y, por ende, una relación forzosa en la que los aventajados parasiten a quienes estén en desventaja: las *asimetrías de información*. Por ahora no me detengo a examinar este asunto que, como veremos, y aunque Gauthier expresamente no se refiera a él, tal vez sea una de las razones por las cuales nuestro autor sostiene su ya mencionada tesis (que en breve se volverá a comentar) de que la moral se hace necesaria como respuesta racional a los fallos del mercado.

aseguraría un mercado perfectamente competitivo a cada agente económico: el que cada uno pueda ser tan libre como Robinson, quien sólo tiene que desplegar una racionalidad paramétrica, dado que no se tiene que enfrentar con ningún otro agente. En la isla de Robinson no habría nadie que pudiera ejercer un control, en su propio beneficio y posiblemente en contra de Robinson, de los términos en los que se daría una interacción con dicho personaje.

Así, para Gauthier, un mercado perfectamente competitivo tendría que asegurar que todos gocen de una libertad como aquella de la que disfruta *Robinson*. Esta idea de libertad, en mi opinión, explica el enorme atractivo moral que tiene para nuestro autor el modelo ofrecido por el *mercado* y, en conexión con éste, el modelo del Robinson solitario que puede actuar libremente dentro de dicho mercado.⁵³ De allí que, para el filósofo canadiense, aquel tipo de situaciones en las cuales los agentes se encuentran en circunstancias completamente opuestas a las de su Robinson libre, es decir, aquellas en las que unos agentes resultan ser explotados por otros, son las que caracterizan a los mercados imperfectos, en los cuales se abre una brecha entre las dos direcciones en las que opera la racionalidad estratégica: el beneficio mutuo y el beneficio netamente individual. Según Gauthier, en ello justamente reside el origen o la razón de ser de la moral: en la necesidad de cerrar esa brecha,⁵⁴ y esto explica por qué los calificativos morales de “bueno” o “malo”, que no tendrían sentido en el contexto de un mercado

⁵³ Sin embargo el lector tal vez podría señalarle a Gauthier que en este caso, al hacer énfasis en la deseabilidad moral de la *libertad* que caracteriza a un agente pensado como un Robinson, por una parte podríamos estar hablando de un *valor* sustantivo, lo cual es algo que, como hemos visto, expresamente había condenado nuestro de autor. Por otra parte, también se podría objetar que un agente solitario, como Robinson, no participa en ningún tipo de interacción ni, por tanto, de mercado; lo cual haría que fuese un tanto trivial hablar de él como de un agente “libre” y, además, asociar esa supuesta libertad a un tipo específico de mercado.

⁵⁴ Con relación a esta tesis de Gauthier, y a riesgo de ser reiterativa, me permito hacer énfasis en su estrecha conexión con un concepto que anima de manera importante su propuesta: la noción de *libertad*, la cual termina por tener un enorme peso en su discurso. Como se ha señalado antes y se volverá a comentar, y aunque aquí el filósofo canadiense parece asumir la postura de un liberalismo igualitario que se preocupa por los perjuicios causados por las desventajas económicas, sin embargo, su concepción de la libertad económica-moral es más propia del tipo de liberalismo con el que realmente simpatiza: el libertario. Para éste, *sólo algunas* de esas desventajas serían relevantes: aquellas que se originan en las imperfecciones del mercado. Las diferencias entre ambas formas de liberalismo explicarían por qué, si se asume una postura cercana al igualitarismo, no tendría mucho sentido hablar de un Robinson solitario como “libre” (tal como se señaló en un pie de página anterior), pues la libertad de la que goce un individuo (o de la que se vea privado) dependería de qué tan *simétrica* sea la estructura de *relaciones* que le vinculan con *otros* agentes. Por consiguiente, para esta forma de liberalismo sería relevante una clase *más amplia* de desventajas, no sólo aquellas debidas a los fallos del mercado, que son las que preocupan a Gauthier y al liberalismo de corte libertario.

perfectamente competitivo, en cambio *sí* se justifican en los mercados que acusan imperfecciones.

We assess outcomes as right or wrong when, and only when, maximizing one's utility given the actions of others would fail to maximize it given the utilities of others (p. 93).

De este modo Gauthier asume que se ha logrado demostrar concluyentemente su tesis de que aquellos constreñimientos morales que se harían innecesarios en un hipotético mercado perfecto, se vuelven imprescindibles y deseables allí donde no se dan las condiciones del mercado ideal, esto es, en aquellas circunstancias en las que los mercados presentan fallos a causa de los cuales unos agentes son explotados por otros. A la luz de lo anterior, creo que llegados a este punto surgen algunas preguntas. En las mencionadas circunstancias en las que la moral soluciona los problemas originados por las imperfecciones del mercado, ¿qué tipo de *agente* es el que actúa, puesto que es claro que no podría tratarse de un Robinson solitario y absolutamente libre? En los mercados reales, con fallos, es decir, tratándose de situaciones en las que se hace necesaria la moral, ¿coincidirían, entonces, el agente *moral* con el agente *económico*, esto es, estaríamos hablando de uno y el mismo agente? La respuesta de Gauthier a esta última cuestión es positiva, si bien el autor se hace cargo de las críticas que se le pueden presentar una vez que ha señalado que el origen y la justificación del ejercicio de la agencia *moral* se encuentran en la necesidad de superar los *fallos* del ámbito *económico*. Pues esta tesis, como arriba mencioné, se halla en tensión con las apologías que el filósofo canadiense hace, tanto de la sociedad de mercado, como del mercado mismo, tomados como ejemplos de moralidad. Veamos, pues, cómo enfrenta el autor algunas de estas críticas.

Gauthier trata de responder a dos objeciones que podrían presentarle quienes intenten descalificar su modelo de agente pensado como un participante del mercado, quien sería lo que el autor también llama nuestro “yo-del-mercado” (*market-self*), y al que luego se referirá utilizando la expresión más conocida: *homo oeconomicus*. Las críticas a las que se enfrenta el filósofo canadiense apuntan, por un lado, a lo distorsionante o poco realista que puede parecer a los ojos de algunos, según él mismo lo reconoce, la figura de un sujeto *moral* pensado bajo el modelo de un agente *económico* y *autointeresado*, tal y como Gauthier lo propone. Por otro lado, habría

quienes señalarían lo antipático y nada “moral” que este personaje pueda resultarnos, a pesar de que, como el propio autor también parece aceptarlo, podría ser más real de lo que desearíamos, ya que admitiría ser visto como un indeseable pero también innegable producto de la forma de socialización que caracterizaría a nuestra sociedad capitalista (pp. 100-101). Esta última acabaría, según las objeciones a las que intenta responder Gauthier, moldeando a sus miembros de una manera tal que éstos se parezcan cada vez más a sus *market-selves*: unos individuos definidos simplemente por sus funciones de utilidad, sus derechos de propiedad, su dotación natural (*natural endowment*) pero, sobre todo, su *autointerés* y su esencial *no-concernimiento* por la suerte de sus congéneres.⁵⁵

No obstante lo dicho, nuestro autor contradice enérgicamente a quienes sospechen de la imparcialidad que reinaría en este mercado perfecto o, peor aun, parezcan atribuirle un *mal moral* esencial que se traduciría, como acaba de mencionarse, en que el sistema moldee a sus miembros y no precisamente para que desarrollen actitudes cooperativas ni capacidad para la autonomía. Según esta crítica, los agentes terminarían por tener las preferencias y los factores de producción que de hecho tienen, con lo cual el mercado se reproduciría a sí mismo, mientras que quienes en él transan sólo serían los medios del sistema. Así, este último, mas no los sujetos, sería lo que realmente importa, mientras que el valor que puedan tener los agentes quedaría reducido a muy poco: tan sólo al que tendrían las piezas de una gran máquina, incluso si ellos no son concientes de esto y actúan creyendo que sus elecciones se orientan a maximizar su utilidad individual. Frente a este problema que se le atribuye al mercado, como sistema ajeno e incluso hostil a la moral y a la autonomía de los individuos, nuestro autor advierte que él realmente *no* cree que nosotros, como agentes morales, debemos identificarnos con nuestro yo-del-mercado. Su apelación al modelo de *Robinson Crusoe* se debe, según Gauthier, simplemente a que todos, en parte, somos y aspiramos a ser un poco robinsones. Pues en esta figura se aprecia una idea de *libertad* que resulta ser muy atractiva y, sobre todo, que es de la mayor importancia para entender, tanto la necesidad

⁵⁵ En este caso, pienso que también se le podría preguntar a Gauthier: los agentes así pensados (como yoes-del-mercado) ¿conformarían el mercado perfecto que, en razón de su estructura, y según lo ya dicho por el autor, haría innecesario que los individuos restrinjan su conducta maximizadora siguiendo códigos morales o, por el contrario, estaríamos ante un mercado con fallos y en el que, además, los agentes se caracterizarían por su autointerés, su no concernimiento mutuo y su irrefrenable conducta maximizadora, haciéndose, por tanto, necesaria y deseable la intervención de la moral?

de la moral, como la del mercado. Se trata de aquella libertad que, según lo afirma nuestro autor, todos deberíamos o quisiéramos tener con el fin de poder dirigir nosotros mismos nuestros talentos y esfuerzos, así como nuestra “dotación natural” (la cual incluye nuestras capacidades mentales y físicas) al servicio de *nuestras* preferencias, y no al de las preferencias de otros agentes. Es decir, a la satisfacción de nuestros intereses y no de los intereses ajenos ni, mucho menos, de unos supuestos intereses colectivos.

We do want to argue that each person is defined in part by an appropriately-based factor endowment . . . this is presupposed even in our account of Robinson Crusoe and the freedom she enjoys to direct her capacities to the service of her preferences (p. 99).

A estas alturas de su argumentación (pp. 98 y 99), Gauthier no ofrece una solución al problema que plantea la perturbadora idea de un mercado que moldearía a sus agentes y que, por tanto, estaría lejos de ser el reino de la autonomía y la moralidad. No obstante, el autor anuncia que responderá a esta objeción, si bien sólo lo hará hasta el final de su texto, cuando intente corregir su figura del *homo oeconomicus* mediante los aportes de su idea de un “individuo liberal”. Por lo pronto, el canadiense intenta mostrar la necesidad de asumir su presupuesto del *no-concernimiento* mutuo, propio de los agentes del mercado. Para ello incluso apela a la autoridad de un filósofo moral que, como Kant, creo que difícilmente podría ser asociado a estas ideas del mercado y del egoísmo de quienes en él participan. Empero, según Gauthier, su propuesta es bastante cercana a un tema kantiano: la tesis de que debemos aplicar las restricciones morales independientemente de los gustos e intereses de aquellos por quienes nos preocupamos. La persona podría estar interesada en la suerte de otros sujetos, o sentir algún vínculo de afecto o de lealtad por algunos de ellos, pero este interés y estos vínculos no serían pertinentes para los mandatos morales. Es más, desde una postura kantiana, estos últimos deberían aplicarse y ser obligatorios al margen de cualquier interés en el bienestar de otras personas o de los lazos afectivos que se pueda tener con ellas. Estamos, pues, según el autor de *La moral por acuerdo*, ante una idea kantiana que dice querer rescatar, al hacer énfasis en su tesis del *no-concernimiento* mutuo que debe atribuirse a los agentes que transan en un mercado.

En mi opinión, dos problemas que el lector podría ver acá son, en primer término, que si se tratase de un *mercado* en el que *se haga necesaria la moral*, entonces no sería fácil ver por qué razón el filósofo canadiense apela a una supuesta idea kantiana de moralidad como *no-concernimiento* mutuo si, precisamente, dicho *no-concernimiento*, según lo ha afirmado el mismo Gauthier, es característico de aquellos agentes que operan en una zona *libre de moral*. En segundo término, creo que en este punto se le podría recordar a nuestro autor que un agente propiamente *kantiano* aplicaría la ley moral, no sólo independientemente de los intereses y gustos de *otros* individuos, sino también al margen de los *suyos* propios. Lo cual incluiría aquello que, usando el lenguaje de los teóricos de la decisión y de Gauthier mismo, conformaría el sistema de *preferencias* del agente. Si un sujeto racional, tal y como lo entiende el filósofo canadiense, no debería decidir al margen de estas últimas, es obvio que, por el contrario, la *moral kantiana* sería ajena a la centralidad que tienen, para Gauthier y para la teoría de la elección racional, dichas preferencias y la búsqueda –por definición, autointeresada– de la satisfacción de éstas. De allí que acaso podría considerarse insólito el acudir a la idea *kantiana* de agencia *moral*, si se está intentando defender el *autointerés* y el *no-concernimiento* (por sus congéneres) como las notas fundamentales del agente modélico que propone el autor de *La moral por acuerdo*.

Tal vez previendo estas objeciones, el filósofo canadiense intenta explicar con mayor detenimiento su problemático supuesto del *no-concernimiento* mutuo, trazando una pintura bastante ilustrativa de su modelo de *agente moral* que, por ahora, en mi opinión, no se ve claramente si se halla completamente separado del *yo-del-mercado* o si se entrelaza con éste. Gauthier acude a esta pintura buscando sortear la innegable dificultad que plantea el describir a los agentes *morales* como agentes *económicos*, en el sentido de no estar concernidos los unos por los otros, y atendiendo cada individuo únicamente a sus propios fines. Pues es claro que resulta difícil aceptar esta descripción por parte de quienes, como ya se ha mencionado y lo reconoce nuestro autor, o bien la consideren poco representativa de cómo somos los agentes morales en la vida real, o bien, por el contrario, la vean como el corazón de la caricatura de una naturaleza humana detestable y poco proclive a la moral, por más que, para algunos, esa caricatura nos recuerde –y no precisamente por irreal– mucho de aquello que más nos pueda preocupar y/o avergonzar de nuestra propia sociedad. La respuesta de Gauthier a estas

críticas es que su idea del no-concernimiento mutuo simplemente da cuenta de un hecho constatable por el mero sentido común: los seres humanos solemos preocuparnos fundamentalmente por nosotros mismos, nuestros amigos y parientes. En cambio, al decir de nuestro autor, nos sentimos poco o nada concernidos por aquellos otros agentes que no pertenecen a nuestro círculo de afectos, salvo en circunstancias excepcionales en las que alguien ajeno a ese círculo se encuentre en una situación de extremo peligro y tengamos la posibilidad de ayudarle. Por tanto, el filósofo canadiense insiste en que su pintura del hombre común y corriente como un ser básicamente egoísta, no resulta para nada falsa ni repulsiva. De allí que su supuesto del no-concernimiento mutuo no deba escandalizarnos, ni hacernos ver un monstruo moral en el modelo de agente que Gauthier nos presenta. Por el contrario, según él, tendríamos, por demás, que apreciar con claridad que dicho supuesto constituye un elemento fundamental, tanto de la lógica con la que funcionan las interacciones en el *mercado*, como de aquella que gobierna en general *todas* las interacciones humanas.

The assumption of mutual unconcern may be criticized because it is thought to be generally false, or because true of false it is held to reflect an unduly nasty view of human nature, destructive not only of morality, but of the ties that maintain any human society. But such criticism would misunderstand the role of the assumption. Of course persons exhibit concern for others, but their concern is usually and quite properly particular and partial. It is neither unrealistic nor pessimistic to suppose that beyond the ties of blood and friendship which are necessarily limited in their scope, human beings exhibit little positive fellow-feelings. Where personal relationships cease only a weak negative concern remains, manifested itself perhaps in a general willingness to refrain from force and fraud if others do likewise, and in a particular willingness to offer assistance in extreme situations . . . But this limited concern is fully compatible with the view that each person should look after herself in the ordinary affairs of life with a helping hand to, and from, friends and kin (pp. 100-101).

3.3 El carácter ejemplar de la sociedad de mercado. La fusión entre agencia moral y agencia económica

Es posible que nuestro autor haya caído en cuenta (si bien no lo dice expresamente) de una objeción que podría hacérsele a este (su) argumento del “sentido común” que él le atribuye al agente moral no-concernido. A saber, que puede señalarse un hecho que en absoluto resulta insólito en el ámbito de las interacciones humanas: que *también* en ocasiones nos preocupamos por la suerte de *otros* sujetos a los cuales *no* nos

ata un previo vínculo de afecto. Es más: ni dicha preocupación ni, por el contrario, la falta de ella, necesariamente tienen que considerarse como justificadas en razón un tal vínculo, ni tampoco únicamente por el “tamaño” del peligro que corre quien demanda auxilio, si bien cabe indicarse igualmente que, como de algún modo lo reconoce Gauthier, entre mayor sea dicho peligro y menos riesgos se corran al prestar la ayuda, solemos ser más proclives a condenar la insolidaridad de quien se niegue a ayudar. El punto es que si ocurre que alguien se abstiene de socorrer a otra persona, a pesar de que puede hacerlo y sin que en ello “se juegue la vida”, e independientemente de que el sujeto en peligro sea –o no– uno de los “suyos” o de los “nuestros”, pertenezca –o no– al círculo de afectos de quien es objeto de la demanda de socorro, el caso es que *no* encontramos excusable su conducta a menos que nos ofrezca una explicación realmente convincente. Y es muy probable que dicha explicación no nos satisfaga si sólo apela al hecho de que ese “otro” que solicita ayuda es, para quien se niega a prestársela, un simple “extraño” ajeno a sus lazos de parentesco/amistad. Tampoco creo que nos parezca muy convincente el argumento de que la situación no reviste “tanto” peligro como para que se justifique el socorrerle (v.g. sólo se va a resbalar “un poco”, no a romperse la crisma); o que el agente a quien se le ha solicitado ayuda tendría que incomodarse o pagar ciertos costes (podría llegar tarde a una cita, etc.). De hecho, algunos autores contemporáneos ven en nuestras *reacciones de reproche* ante ciertas muestras de insolidaridad (y no sólo tratándose de “graves” peligros para quien demande la ayuda), una forma de experiencia que para nada resulta extraña en nuestra vida en sociedad y que señalaría el ámbito de aquello que estaríamos dispuestos a llamar “lo moral”.⁵⁶

Creo que un ejemplo bastante ilustrativo de esto que aquí se intenta mostrar podría verse en el intento de E. Tugendhat por determinar qué es eso que, por lo menos a partir de la modernidad, solemos llamar “lo moral”. Su respuesta indica que las demandas y expectativas mutuas que según él parecen hacer parte del campo moral, se presentarían en (o atravesarían) un amplísimo conjunto de praxis que no se reducen a la posibilidad de establecer vínculos de afecto. Sobre todo, dicha “amplitud” significa que en franco contraste con otras formas de praxis más *reducidas* –v.g. un juego específico, como el fútbol, cuyas reglas sólo rigen en aquellos escenarios donde se jueguen partidos

⁵⁶ Cf. Strawson (1995), y el uso que, de los aportes de dicho filósofo, hace J. Habermas (1994a).

de fútbol—, según Tugendhat aquello que llamamos “moral” atraviesa o “permea” unos terrenos mucho *más incluyentes* en el que pueden situarse diversas formas de praxis —entre muchas otras, también los partidos de fútbol—. ⁵⁷ Pienso que esto podría autorizar que en este punto del análisis de su propuesta se le señale de nuevo a Gauthier que estaría restringiendo excesivamente el ámbito de aplicación de “lo moral”, pero esta vez de manera más explícita por la vía de extender acaso demasiado el campo en el que regirían las reglas del juego económico, y los supuestos de los que tiene que partir quien analice esa porción específica de la vida social. Por ello cabría aquí la sospecha de que este tipo de análisis y de supuestos aplicables a lo estrictamente económico, y que resultan aceptables e incluso imprescindibles para modelar lo que ocurre dentro de dicho ámbito de interacciones, tal vez no sean los más apropiados para modelar el ámbito de *todas* o de muchas de las interacciones humanas, incluyendo de manera especial a la moralidad.

Sin embargo, Gauthier insiste en que la asunción de su idea de unos seres no-concernidos los unos por los otros parece estar de acuerdo, no sólo con nuestras intuiciones de sentido común, en cuanto a cómo suelen comportarse los agentes humanos en los contextos corrientes de la interacción social en general, sino también con el tipo específico de cooperación que constituye el mercado. En este último los agentes, en tanto que no-concernidos los unos por los otros, no deben ser pensados como teniendo lazos de amistad ni de parentesco y, por ende, sus funciones de utilidad han de tomarse como independientes. No obstante, en mi opinión algo que sorprende de esta situación tal y como la presenta el autor es que, según él, se trata de un estadio *moralmente superior* a aquellos otros escenarios en los cuales normalmente se piensa a las personas que mantienen esos vínculos de afecto o, por lo menos, de solidaridad, ausentes del esquema del mercado. Pues para el filósofo canadiense los *yoes-del-mercado*, a diferencia de quienes están relacionados afectivamente entre sí, pueden

⁵⁷ Esto explica para Tugendhat por qué quienes toman parte en ese tipo de interacciones que conformarían lo moralmente relevante —esto es, ese campo relacional que parece ser más incluyente o más amplio que las diversas formas de praxis que lo moral “permea” o “atraviesa”—, se sentirían impelidos como agentes morales, bien por los demás sujetos, bien por sí mismos, a cumplir con tales normas y expectativas mutuas; y ello incluso allende la posibilidad de “entrar/salir” de los escenarios específicos donde se desarrollen prácticas más *restringidas*, tales como los partidos de fútbol. De allí que, según lo dicho por el autor alemán, no sea tan fácil “salirse” de una praxis tan amplia o transversal como la moral, ni justificar el hacerlo, como en cambio podría suceder con el hecho de “salirse” de un partido de fútbol y, por ende, declararse libre de las ataduras del reglamento de dicho juego concreto cuando uno no está tomando parte en él. Cf. Tugendhat, E. (1988, pp. 84-89; 172-195).

transar y cooperar con aquellos otros agentes con los cuales no desarrollen tales vínculos. Según Gauthier, esto contrasta con lo que suele ocurrir en las sociedades pre-capitalistas o de mercados no desarrollados, en las cuales existe la tendencia a hacer negocios sólo con aquellos en quienes se confía en razón de los lazos de parentesco o amistad. Esta situación “primitiva” es superada, afirma el autor, por la sociedad de mercado, y ello se constituye en una muestra de la superioridad moral de esta última.⁵⁸ Quienes viven en ella pueden escoger con quiénes se relacionan, tanto para hacer negocios y llevar a cabo sus actividades económicas, como para construir sus relaciones afectivas. En un contexto así, estas últimas tampoco les son impuestas a los agentes como, en cambio, sí les ocurre, nos dice Gauthier, a quienes viven en sociedades pre-capitalistas o no-capitalistas. Para él, confianza y libertad en los negocios, así como liberación de los lazos de la tribu, son las ganancias innegables de la moralidad propia de la sociedad de mercado; ganancias que se traducen en el resultado que produciría la mano invisible: cada uno, persiguiendo su propio interés, aporta al de todos.

The superiority of market society over its predecessors and rivals is manifested in its capacity to overcome this limitation and direct mutual unconcern to mutual benefit. If human interaction is structured by the conditions of perfect competition, then no bond is required among those engaged in it, save those bonds that they freely create as each pursues his own gain. The impersonality of market society, which has been the object of wide criticism and at the root of charges of anomic and alienation in modern life, is instead the basis of the fundamental liberation it affords. Men and women are freed from the need to establish more particular bonds . . . (pp. 101-102).

Llegados a este punto, creo que el lector puede apreciar con más claridad algo que arriba ya se mencionó: un preocupante desliz por parte de Gauthier, quien parece haber pasado sin mediaciones desde la asunción de una metáfora, la del mercado perfecto y sus *market-selves* —tomadas como *ficciones* necesarias para modelar a la sociedad en tanto que sistema de cooperación, y a los agentes que en ella participan—, a lo que ahora puede ser visto como el elogio de un tipo *real-histórico* de sociedad, la cual debe ser pensada como el ejemplo a seguir por parte de cualquier otra (actual o futura). A mi entender no queda claro si, por una parte, estamos hablando en términos de ficciones o metáforas útiles para modelar ciertas realidades a un nivel meramente

⁵⁸ Al respecto véase lo que afirman, en este mismo sentido, autores como Friedman (1966 y 1980); Hayek (1978, 1991 y 1998); y Fukuyama (1996). Más adelante se mencionará el tema de la *confianza* y se verá que ésta puede también ser entendida de otras formas distintas a como la entiende Fukuyama.

conceptual y normativo o si, por la otra, estamos al nivel de hechos verificables. Tampoco creo que quede claro el alcance que finalmente quisiera darle Gauthier a sus metáforas ni, por consiguiente, cómo concebir la relación y las posibles distancias que habría (en el contexto de su discurso) entre lo fáctico y lo modélico. Pues, de un lado, como hemos visto nuestro autor se defiende en contra de quienes querrían criticarle por tomarse demasiado en serio sus ficciones o metáforas y creer, por tanto, que efectivamente así de monstruosos y de egoístas es como *somos* los seres humanos, y así de insolidarias son las sociedades que *de hecho* establecemos; esto es, meros *market-selves* interactuando en meros mercados. Pero, de otro lado, ahora resulta que Gauthier nos dice que el mercado es la *mejor* expresión de lo que somos y de lo que *deberíamos aspirar a ser* todos los agentes humanos, pues en él se da, como un *hecho* y no sólo como una *idea regulativa*, la liberación de los lazos de la tribu, tanto para hacer negocios como para establecer sin coerción nuestras lealtades y afectos personales.

Empero el autor intenta, una vez más, adelantarse a este tipo de críticas y entra a comentar el tema de los peligros que ofrece esta suerte de liberación operada por el mercado en aquel tipo de sociedades en las que no está del todo consolidado el capitalismo; peligros en los que se combinan, según Gauthier, los fallos de ambas formas de sociedad: la que se encuentra en un estadio previo al capitalismo y aquella otra en la que dicho sistema se encuentra más desarrollado. Así, nuestro autor afirma que en los escenarios sociales cuyo capitalismo aún es incipiente, por un lado, suele entorpecerse el cumplimiento de las restricciones morales y políticas que pondrían límites a las ingerencias de las lealtades personales o de tribu en los asuntos públicos y, de otra parte, esta preocupante situación se agrava aun más, ya que a ella se suma la tendencia a que, al mismo tiempo, la vida de los afectos se contamine por causa de la intromisión de intereses meramente económicos. Sin embargo, no ha de perderse de vista que estos resultados adversos sólo se dan, según Gauthier, en sociedades que están a mitad de camino de su desarrollo capitalista, no en aquellas otras en donde ya se ha consolidado plenamente el sistema de mercado. Este último tipo de sociedad sigue siendo, a su entender, ejemplar. Allí las restricciones morales no sólo son perfectamente compatibles con el no-concernimiento mutuo que hay entre los ciudadanos sino que, además, se hacen necesarias a causa de ese mismo no-concernimiento, y dada la estructura de relaciones que se establece entre los agentes que interactúan de este modo.

En este tipo de escenarios sociales la *moralidad* no tiene un fundamento *afectivo* sino que se respeta en razón de la búsqueda del *interés individual*, esto es, posee una base propiamente *racional*.

Morality, as a system of rationally required constraints, is possible if the constraints are generated simply by the understanding that they make possible the more effective realization of one's interests, the greater fulfilment of one's preferences, whatever one's interests or preferences may be. One is not then able to escape morality by professing a lack of moral feeling or concern, or a lack of some other particular interest or attitude, because morality assumes not such affective basis (pp. 102-103).

En cuanto a esta idea de un fundamento racional y no afectivo de la moral, Gauthier nuevamente manifiesta hallarle la razón a Kant y, por el contrario, no compartir una posición como la de Hume. No obstante, una vez más pienso que se le podría objetar al autor canadiense que resulta un tanto difícil emparentar a Kant con una defensa del sistema de mercado. Entre otros motivos porque no creo que el filósofo alemán estuviera postulando su teoría moral como aplicable a la sociedad humana modelada en términos de vínculos *económicos* los cuales, a los ojos de Kant, tal vez serían vistos como relaciones de *mutua instrumentalización*. Si para él esto último constituye aquello a lo que precisamente habría que poner *límites* mediante la *moral*, entonces obtendríamos de Kant una pintura poco conciliable con la apología que hace Gauthier de la sociedad de *mercado* como *moralmente ejemplar*. En mi opinión el filósofo alemán estaba pensando, en franco contraste con los lineamientos de *La moral por acuerdo*, en términos de un *respeto moral* que, si bien no obedece a simpatías personales, como correctamente señala Gauthier, tampoco viene determinado por la lógica económica de la mutua instrumentalización. Por tanto, no sería acertado atribuirle un carácter *moral-kantiano* a la persecución de los fines individuales, llevada a cabo por los agentes que no están concernidos por sus congéneres. Al contrario: antes que obedecer a aquello a lo que Gauthier llama “no-concernimiento mutuo”, los mandatos morales, tal y como se piensan desde Kant, más bien constituirían los límites que el respeto impone a quienes, persiguiendo sus propios intereses, desatienden los intereses de las demás personas.

Otra objeción que a estas alturas pienso que se hace aun más justificada apuntaría a que Gauthier primero ha atribuido al mercado perfecto una ausencia –por

innecesaria e injustificada— de la moral y, sin embargo, ahora resulta que nos dice que esta última sólo podría surgir en una sociedad de mercado perfecto, ya que en ella sus miembros, al no estar concernidos los unos por los otros, están unidos por relaciones que no son afectivas sino meramente racionales. Con lo cual ocurre que cada uno procura su propio interés y, por tanto, sólo respeta las reglas del juego como el vehículo necesario para el logro de dicho interés. De manera que nos encontramos con la incómoda conclusión de que no queda claro, finalmente, si se está hablando de un mercado perfectamente competitivo como zona *libre de moral* o si, por el contrario, se está señalando que dicho mercado es el único contexto en el que *puede surgir una moral* racionalmente justificada. Esta confusión tal vez se deba, insisto, al hecho de que Gauthier parece tomarse demasiado en serio sus modelos del mercado y de sus *market-selves*; tan en serio que por momentos dejan de ser metáforas o esbozos orientativos y pasan a ser el “ejemplo real” a seguir. Se trata del ejemplo constituido por aquello que, usando una suerte de jerga hegeliana, podría llamarse lo “real-racional”: la realidad ejemplar encarnada en la moderna sociedad de mercado.⁵⁹ Esta última, y sin que medie, repito, alguna explicación que lo haga comprensible, súbitamente deja de ser el *locus* de mercados meramente imperfectos, a pesar de ser real-histórica y aunque parezca necesitar restricciones morales, a la vista de los innegables fallos e injusticias —moralmente preocupantes— que se le podrían señalar.

Uno de estos fallos que, por cierto, Gauthier no menciona expresamente, viene dado por las asimetrías de información. Si bien, como digo, nuestro autor no se refiere a este asunto de manera explícita, es posible que lo tuviera en mente al postular su tesis de la moral como la solución racional a las imperfecciones del mercado. De este modo, si se piensa en una hipotética situación en la cual ya se han solucionado los problemas ocasionados por tales imperfecciones, es de suponerse que, para el filósofo canadiense, el sistema de interacciones económicas entraría a funcionar tal y como lo haría un mercado perfectamente competitivo, lo cual exigiría la erradicación de todo tipo de transacciones forzosas, incluyendo aquellas que se originan en las *asimetrías de información*. Esto es, situaciones caracterizadas por una desventaja y, con ello, por una pérdida de libertad para quien “sabe menos”, frente a quien transa con ella/él “sabiendo

⁵⁹ Esta idea de una defensa de la moderna sociedad de mercado como lo real-racional, en el sentido hegeliano de la expresión, aparece lucidamente expuesta por F. Ovejero (1994a y 1994b).

más”; v.g. tratándose de un agente que compra un producto o un servicio conociendo muy someramente las características de lo que compra, frente a quien le vende dicho producto o servicio sabiendo a ciencia cierta qué es eso que le está vendiendo a su cliente. Esta posición de ventaja del vendedor sobre el comprador, a causa de su mayor acopio de información, le permitiría al primero “parasitar” y explotar al segundo, y en ello creo que Gauthier estaría de acuerdo. De allí que seguramente su tesis de la moral como el remedio para las imperfecciones del mercado, incluiría la corrección de las asimetrías de información.

Pienso que nuestro autor tampoco objetaría que, a partir de esta tesis suya, se obtuviera así mismo la conclusión de que, si ocurriese que en el mercado se vieran favorecidos los intereses del explotador, en contra de quien es objeto de esta particular forma de explotación producida por las asimetrías de información, entonces en dicho mercado ocurriría que algunos agentes *no* podrían elegir *libremente*, al carecer de aquello que Gauthier llama preferencias “consideradas”. Pues la falta de *cualificación de las preferencias* de quien sabe menos, sería la causa de que se viese forzado a elegir en contra de sus propios intereses, incluso sin estar conciente de ello. Estaríamos por tanto ante una relación coercitiva originada en la *no-cualificación de las preferencias* de la parte en desventaja y en ello, repito, nuestro autor seguramente concordaría. No obstante no estaría muy segura de que Gauthier también estuviese dispuesto a conceder que, si lo anterior demuestra la necesidad de que los agentes tengan preferencias “más racionales” *i.e.* más cualificadas, con el fin de que se vean libres de una ignorancia que les pondría en desventaja frente a otros agentes, entonces esto demostraría que la *racionalidad de las preferencias* sí depende, después de todo, de los *contenidos* de tales preferencias o, mejor, de la *cualificación* de tales contenidos. Esto último, y tal vez a despecho de lo que anteriormente ha sostenido Gauthier, requeriría el apelar a *referentes normativos* para juzgar tales preferencias, no bastando por consiguiente con la asepsia valorativa en la que querría instalarse nuestro autor. Es decir, éste tendría que reconocer que las preferencias racionales deben cumplir con estándares normativos mucho más exigentes que la mera consistencia lógica que tendría que caracterizar, según los cánones de la teoría de la decisión, a los sistemas de preferencias.

Amén de lo anterior, creo que también se le podría recordar a Gauthier que, si como algunos autores insisten, muchas de las mencionadas asimetrías de información

son insalvables incluso en un mercado teóricamente perfecto –v.g. la asimetría que separa a un paciente de su médico–, esto podría llevar a pensar que, para garantizar relaciones no-forzosas dentro del mercado, resultarían insuficientes los mecanismos que el propio mercado genere. Aunque seguramente Gauthier estaría de acuerdo con esto último, tal vez no comparta la idea de que casos como el del médico y su paciente mostrarían la necesidad de que en algunas o muchas situaciones normales de las interacciones humanas, interacciones que también admiten ser vistas desde la óptica económica, un agente no tiene más remedio que *confiar* en aquel otro agente con el que transa. En este caso concreto, el paciente no tendría otra salida que la de confiar en su médico puesto que su desconocimiento del servicio que le vende el facultativo resulta prácticamente insalvable. Esto haría caer en la cuenta de la necesidad de *otras instituciones* adicionales al mercado mismo, instituciones que se generan en diversos campos de interacción, como el político, el legal-judicial, pero también en un ámbito que a lo mejor resulte no encajar en el esquema de Gauthier: las *costumbres* o la moralidad entendida al modo “tradicional”; es decir, definida en aquellos términos que nuestro autor tanto descalifica.

Las mencionadas instituciones, más los “vasos comunicantes” que algunos creen que las unen, garantizarían que las interacciones entre las personas, así como las relaciones entre éstas y las propias instituciones –incluyendo también al mercado mismo–, “fluyeran”. Ello, dado que así se permitiría que unos agentes pudiesen *confiar* en los otros sin necesidad de estar en permanente alerta frente al peligro de ser explotados, alerta que implicaría un aumento en los costos de transacción, y una constante parálisis de las decisiones individuales que sean económicamente relevantes.⁶⁰

⁶⁰ Cf. K. Arrow (1996) y véase, en especial, su polémica con M. Friedman (1996)), quien descalifica el que se apele a temas ajenos a la dinámica del mercado, tales como una supuesta “responsabilidad social” (i.e. ética) de las organizaciones empresariales, cuando se tiene el amparo de un aparato judicial eficiente y de un estado bien organizado. Este último, por tanto, al decir de Friedman y en contra de lo que piensa Arrow, no debe interferir en el libre despliegue del mercado, ni tampoco los agentes económicos deben restringirse en su libre actividad, salvo para respetar la ley. Por el contrario, para Arrow, no basta con la ley y, por ello, es necesario suponer una base de confianza que sería impensable si los agentes económicos no compartieran y dieran por sentado que los otros también comparten ciertos códigos morales, para cuyo cumplimiento saldría muy caro el apelar constantemente al aparato judicial, entre otras razones, porque querellar es muy costoso y porque, una vez que se causan ciertos daños, la reparación es, o bien imposible, o bien muy limitada. Como puede apreciarse, esta idea de confianza basada en códigos morales no concuerda del todo con aquella que defendería un autor como Fukuyama (1996), quien estará mucho más cerca de una posición como la de Friedman. Véase también; la importancia que autores como A. Sen (1992) y D. North (1994) le conceden a esas *otras instituciones* que, aunque no obedecen a los mecanismos del mercado, interactúan entre sí y con dichos mecanismos

El punto al que voy es que tal vez estas dos observaciones que acaban de hacerse, es decir, 1) La necesidad de *referentes normativos* que *cualifiquen* el *contenido* de las preferencias, y 2) La necesidad de *instituciones* que, como los códigos de las *morales* “tradicionales” y aquellos que se legitimarían desde discursos de carácter más político e incluso filosófico, se entrelacen con las reglas de juego del mercado y, por tanto, *no se restrinjan a lo meramente económico y legal*. Con esto se apunta de nuevo a las razones para dudar de una tesis fundamental para Gauthier: la del *mercado* como modelo de lo *moral*; o la de los agentes *económicos* como modelos de agencia *moral*.

En mi opinión, esta objeción no encontraría una respuesta satisfactoria en el texto de Gauthier. No insistiré por ahora más en ella, pues creo que ya el lector se habrá hecho cargo de las dificultades que entraña el proyecto, a mi entender difícil de lograr, de modelar las relaciones entre agentes *morales* en términos de interacciones entre agentes *económicos* y, en conexión con ello, el postular al *mercado* (que, repito, a estas alturas ya no se sabe si se trata del *mercado perfecto* o si es el *mercado real*) como modelo de sociedad *moralmente ejemplar*. A lo cual habría que agregar que todo ello se ha argumentado o elaborado con el fin de proveer a la *moralidad* de un sustento *racional*, explicando el origen o la necesidad que se tiene de la primera en tanto que respuesta racional a los *fallos del mercado*; un mercado que paradójicamente, repito, se supone que provee un modelo ejemplar –y moral– de cooperación social. Visto lo anterior acaso podría intuirse que, tal y como arriba se mencionó, detrás de todas estas dificultades que se han mostrado esté, finalmente, el interés que creo que es clave para Gauthier: salvar por encima de todo a su modelo de *agente* pensado como un *yo-del-mercado*; como un Robinson *libre y egoísta*, cuyos únicos lazos con los demás agentes –también robinsones– son sólo los exigidos por la viabilidad del sistema económico de cooperación.

Tal vez por esto último el autor intenta contestar a la objeción que le plantearían quienes pudieran estar preocupados por los *efectos moralmente perniciosos* que tendría, en quienes viven en sociedades capitalistas, este *ethos egoísta*, propio del mercado y de sus *market-selves*. Ante este problema Gauthier les advierte una vez más a quienes teman estos resultados éticamente indeseables de la socialización para la vida en

económicos para hacer posible, no sólo la economía sino, en general, la vida social y las interacciones humanas en sus diversas facetas.

contextos capitalistas que, por una parte, la pintura de sus *homini economici* no es tan moralmente tenebrosa como algunos se empeñarían en mostrar. En segundo término, anuncia que él *no* va adoptar del todo el modelo del *homo oeconomicus*. Antes bien, dice que posteriormente rechazará buena parte de las características de esta figura y, al final de su libro, incluso le hará los matices necesarios, aportados por su modelo del “individuo liberal”. A diferencia del yo-del-mercado, este último conserva, anuncia Gauthier, lazos de afecto con sus congéneres, si bien al mismo tiempo y sin que en ello deba verse una contradicción, seguirá formando parte de una sociedad fundada sobre la base de restricciones morales puramente *racionales i.e.* basadas en el *autointerés*.

A society that frees individuals from the need to develop affective bonds in order to engage in social interaction may encourage its members to conceive themselves as unrelated individual atoms lacking the potential for developing genuinely affective ties. Market man is a maximizer of *non-tuistically* based utilities; our account, it may seem gives market-man a morality but denies him any other-directed concern. The particular affective ties that we acknowledged at the beginning [of this subsection] may be only *the residue of an image of human beings that we have rejected* (p. 104; las cursivas son mías).

Esta es la primera ocasión en la que nuestro autor menciona el término *tuistic*, el cual como veremos, adquirirá gran importancia para lo que resta de la argumentación de Gauthier, y para el que no se me ocurre una traducción adecuada. Por ello me permitiré utilizar el vocablo “tuista”, inexistente en nuestros diccionarios, pero que creo conserva la idea clave para el filósofo canadiense, pues la palabra original designa lo contrario de aquel tipo de valores en virtud de los cuales tenemos en cuenta a *otras* personas. También se trata de valores propios de las tradiciones morales pre-filosóficas, las cuales como ha insistido Gauthier, no constituyen una adecuada base para una moral racional o para una fundamentación racional de la moral.

En el siguiente capítulo me ocuparé de la defensa que hace el autor de su modelo de agente y de sociedad, una vez vistas las fisuras que se le han señalado hasta aquí, sobre todo en lo que se refiere al vacío *moral* que podría achacarse a dicho modelo. Esta defensa consistirá, como veremos, en la propuesta de una *teoría de la justicia* con la cual Gauthier busca precisamente dar cuenta de la legitimidad *moral* de las interacciones entre agentes económicos egoístas. Con esto el filósofo canadiense intentará superar la posible desventaja que podría presentar su “moral por acuerdo” no

sólo frente a las mencionadas formas tradicionales o “tuistas” de moralidad, sino también tratándose de discursos propiamente *filosóficos* que igualmente sean “tuistas” y que, por ende, insistan en entender la agencia moral y los vínculos propios de la vida moral como no-modelables en términos económicos. Veamos, entonces, de qué manera nuestro autor intenta demostrar que su teoría logra hacer frente a los reclamos provenientes de este tipo de posiciones.

Capítulo 2

Cómo unir egoísmo y cooperación: una teoría de la justicia que muestre a la moral como la solución a los fallos del mercado

Frente a críticas como las comentadas en el capítulo anterior, Gauthier intentará llenar el vacío *moral* de su modelo “no-tuista” del agente pensado como un Robinson no-concernido por sus congéneres, el cual, como vimos, al cooperar con otros robinsones tan egoístas como ella/él puede –y sin que en esto nuestro autor vea contradicción alguna–, construir con ellos una sociedad no solamente estable sino, además, *justa*. El filósofo canadiense buscará entonces mostrar que esta forma de cooperación y estos agentes egoístas se constituyen en modelos plausibles de moralidad. Para lo cual desarrolla una teoría de la justicia que respalde su tesis de que la cooperación entre agentes maximizadores, amén de ser la respuesta racional –i.e. basada en el autointerés– a los fallos del mercado, cuenta con el *plus* de ser una respuesta *moral*, dado que les resulta *igual de racional* o de ventajosa a *todos* los implicados. Es decir, se trata de una solución económicamente eficiente en el caso de cada agente y, por tanto, *justa* a los ojos de todos ellos.

Si Gauthier se propone mostrar que dicha solución así mismo debe ser calificada de “moral”, por supuesto que no lo hace entendiendo esto último a la manera de lo que él considera es la forma tradicional de moralidad concebida en clave altruista, esto es, en tanto que renuncia a los propios intereses y deseos, o entendida como la preocupación por los intereses y necesidades del prójimo. Por el contrario, para el autor, la moral definida como cooperación entre maximizadores es fundamentalmente una estrategia de solución racional –o sea, autointeresada– a los problemas causados por los costes que la propia razón le impone a la interacción entre las personas (p. 112). De modo que el reto con el que se encuentra el filósofo canadiense está entonces en demostrar que su concepción de las interacciones humanas en términos de *mercado* también admite verlas como formas de cooperación *justas* o moralmente admisibles. Sin embargo, en lo que sigue se verá con algo más de detalle que la insistencia en modelar

la moral y la justicia en clave *económica*, no dejará de acarrearle serios problemas al tipo de fundamentación *moral* que Gauthier intenta llevar a cabo.

Nuestro autor se refiere al ámbito de la cooperación entre humanos-maximizadores, como el “dominio de la justicia”: “Cooperative interaction is the domain of justice” (p. 113). Lo interesante es que, en claro contraste con otros teóricos de la justicia, en el canadiense encontramos una definición de dicho concepto que, por una parte, se articula en términos puramente *económicos* y en relación con el modelo del mercado. Y, por otro lado, Gauthier advierte que hará uso de un vocabulario al que él se refiere como “moral”, intentando con ello no restringirse a un análisis meramente *político* de la justicia. Con todo lo cual se propone establecer un claro contraste frente a otros aportes, v.g. el de Rawls, quien, como sabemos, busca llevar a cabo una aproximación específicamente *política*, no moral,⁶¹ al asunto de la justicia, si bien el autor de *La moral por acuerdo* reconoce que en el filósofo estadounidense ha encontrado materiales muy importantes para su propia teoría. En cuanto al primer aspecto que señalo, el énfasis en lo económico, repárese por ejemplo en la definición que Gauthier ofrece de la interacción cooperativa —la que se opone a las interacciones forzosas o coercitivas— como aquella que tiene lugar cuando ninguno de los agentes implicados le impone *costes* desproporcionados a ningún otro y, por consiguiente, no se presentan fenómenos tales como la presencia de *parásitos* y *polizones*. Por ahora me centraré en el *primer* aspecto y dejaré momentáneamente de lado el segundo, es decir, el intento de no hacer un tratamiento político sino moral de la justicia, tema sobre el que volveremos hacia el final del siguiente capítulo.

1. La necesidad del contrato social ante la escasez y el autointerés como “circunstancias de la justicia”

Para nuestro autor la cooperación es el tipo *justo* de interacción humana, siendo la justicia “la disposición racional a la cooperación” (p. 113). En este punto Gauthier señala que se encuentra de acuerdo con Hume y con Rawls en cuanto a la idea de que hay unas “circunstancias de la justicia”: unas condiciones bajo las cuales la cooperación

⁶¹ Rawls, J. (1996, pp. 11-23).

humana no solamente es posible sino, además, necesaria.⁶² Advierto de nuevo que el tratamiento que el filósofo canadiense hace de los temas clásicos del contrato social y de la necesidad de la asociación humana se hace en unos términos *económicos* que, en mi opinión, resultan algo distorsionantes tanto para el caso de Rawls, como para el de los padres del contractualismo moderno y, por supuesto, también tratándose del filósofo escocés. Según Gauthier este último acertó al advertir que la necesidad de la cooperación nace de la conjunción de dos factores esenciales de la condición humana: la *escasez* en cuanto a nuestras circunstancias *externas*, y el autointerés (*self-bias*) propio de nuestro *temperamento natural*.⁶³

Esta última característica de los seres humanos que tienden a procurar su propio bienestar, descuidando y en algunas ocasiones malogrando el de sus congéneres, se manifiesta según el autor canadiense en aquello que tanto hemos comentado: el no-concernimiento mutuo que los individuos muestran en sus interacciones en el *mercado*. Si este último fuese un mercado sin fallos permitiría por su misma estructura que dicho no-concernimiento fuese inocuo pero, según Gauthier, Hume muestra correctamente que las circunstancias en las que normalmente vivimos los seres humanos incrementan esta tendencia nuestra a desentendernos del bienestar ajeno y preocuparnos excesivamente por el nuestro, puesto que no podemos escapar a la condición esencial de la *escasez*.⁶⁴ Si ocurriera lo contrario, esto es, si la naturaleza nos proveyera abundantemente de todo lo que necesitaríamos o deseáramos, entonces nos sería más

⁶² Rawls (Ibid, pp. 81-82); Hume, *Investigación sobre los principios de la moral*, Sección 3, parte 1.

⁶³ Cf. Hume (Ibid.). Nótese por parte de Gauthier el recurso a un tema seminal para los clásicos de la economía: la *escasez*, concepto que será de la mayor importancia para la argumentación del filósofo canadiense y sobre el que volveremos reiteradamente. Así mismo nuestro autor insiste en un supuesto psicológico también fundamental en los iniciadores de la economía moderna: el *egoísmo* que habría que atribuir a los agentes que intervendrían en el *mercado*, en cuanto a la conveniencia metodológica de suponer para cada agente económico una *función de utilidad independiente* de las funciones de los demás. Sin embargo, en cuanto al uso que hace Gauthier de ambas nociones, creo que tratándose específicamente de su interpretación de Hume, y de modo análogo a lo que se indicó acerca de su lectura de Smith, podría señalarse que por momentos el autor canadiense parece olvidar que Hume, a pesar de su interés en la economía, en primer lugar no sólo está intentando dar cuenta de las interacciones económicas, sino de las *interacciones humanas en general* y en la posibilidad de explicar cómo es posible la armonía social, a pesar de que las personas manifiestan preocupaciones claramente autointeresadas. En segundo término, a diferencia de Gauthier, Hume no sólo menciona estas tendencias egoístas como parte de la naturaleza humana, sino que igualmente incluye dentro de ésta aquellos sentimientos —empáticos— que muestran que tenemos *también* una capacidad para conmovernos ante los infortunios ajenos. Cf. Hume (Ibid. Sección 2, parte 1, y apéndice 1; y Stroud, B. (1986, pp. 241 y ss.).

⁶⁴ Cf. Hume (Ibid).

fácil ser benevolentes y procurar asimismo el bienestar de los demás.⁶⁵ De este modo no habría polizones, rentistas, parásitos ni explotadores de ningún tipo, ya que todos naturalmente buscaríamos cooperar con los otros y coordinar sus acciones con las nuestras, a fin de obtener bienes comunes. No obstante, afirma el filósofo canadiense, la escasez es insalvable o, más exactamente, no es posible que dejemos de percibir nuestro medio como caracterizado por bienes escasos. Es esta percepción, es esta conciencia (*awareness*) nuestra lo que hace que no sea inocua nuestra natural tendencia al egoísmo, pues al percatarnos de la escasez de recursos y al ver que los demás agentes también quieren apropiarse de éstos, terminamos por percibirles como nuestros hostiles competidores. El resultado de esta percepción es ya un lugar común: nos embarcamos en una suerte de carrera o de pugna de todos contra todos, lo cual termina por imponernos aún más costes de los que acarrearíamos si no estuviéramos tan concientes de la escasez y viéramos de otro modo a nuestros semejantes.

Gauthier así mismo comenta la versión hobbesiana de este tópico: en nuestra conciencia o percepción de la escasez de recursos, aunada a la percepción que tenemos de nuestros semejantes como nuestros hostiles competidores, y agravada por nuestra predisposición a desentendernos de su bienestar al buscar el nuestro, se halla el origen de aquella característica humana que, según Hobbes, no es otra que nuestra preferencia por dominar.⁶⁶ Para este autor, en el estado de naturaleza dicha búsqueda del dominio sobre los otros agentes se debe a que los hombres tienden a pensar que aquel que, mediante el uso de la fuerza, logre imponerse sobre sus congéneres, puede entonces controlar los recursos disponibles y de este modo asegurarlos para sí mismo en circunstancias de escasez. El resultado es el mismo lugar común que acaba de comentarse para el caso de Hume: una guerra latente o explícita de todos contra todos, que se traduce en una situación de precariedad y riesgo para los que así se enfrentan. Igualmente la solución, para el caso de Hobbes, se encuentra en el logro de un *acuerdo* en virtud del cual, y como condición para asegurar la paz, todos los afectados por el indeseable estado de guerra se comprometan a restringirse en el uso de medios violentos. Si antes del pacto cualquier agente estaba justificado a emplear la fuerza en

⁶⁵ Este supuesto bien podría cuestionarse, como puede advertir el lector, si se tiene en cuenta que la abundancia de recursos en las sociedades más opulentas no se traduce necesariamente en una mayor solidaridad.

⁶⁶ Cf. Hobbes (*Leviatán*, Cap. XIII).

contra de los demás, a partir éste, por el contrario, ya puede hablarse propiamente de *justicia*. Esta última, según la lectura que Gauthier hace de Hobbes, descansa precisamente en el mencionado acuerdo. También en el caso del autor del *Leviatán*, como en Hume, esta conciencia nuestra, este darnos cuenta de las circunstancias impuestas por nuestro medio signado por la escasez de recursos y, además, plagado de competidores que se afanan en despojarnos de éstos, no es más que un resultado de nuestro ser-*racionales*. De allí que la *razón* sea igualmente quien tenga que proveernos del remedio a los problemas que ella misma nos causa. Dicho remedio así mismo se obtiene por la vía del “darnos cuenta”: al percibir a los otros ya no sólo como nuestros antagonistas, sino también como nuestros posibles aliados o colaboradores; no sólo como una fuente de conflicto, sino también como una posible fuente de beneficios.

1.1 Algunas dudas sobre la insistencia inicial en el modelo del mercado y un posible desliz de lo político a lo económico

No obstante lo anterior, el filósofo canadiense afirma que Hobbes, si bien acierta al mostrar el rol positivo que cumple la razón al facilitarnos la salida de la situación de guerra mediante el logro de un acuerdo, sin embargo yerra al enfatizar excesivamente los costes que esta misma razón le impone a las relaciones que de modo natural establecemos con nuestros semejantes. Según Gauthier, el clásico británico cree que nos vemos unos a otros básicamente como *competidores* y no tanto como fuentes de beneficio, razón por la cual en este punto el autor de *La moral por acuerdo* dice que prefiere separarse de Hobbes, pues éste presenta una pintura que en su opinión resulta poco aceptable para modelar la manera como los agentes humanos nos percibimos mutuamente. De modo que en lo que se refiere a los problemas para la sociabilidad humana y la solución de éstos, el canadiense se inclina más a pensar este asunto bajo una concepción que le atribuye a Hume y Rawls, dado que para estos dos pensadores lo razonable es que igualmente percibamos a los demás agentes como nuestros posibles cooperadores, si bien nuestra conciencia del mundo no deja de incluir un conocimiento de las imperfecciones y *externalidades* negativas que pueden presentarse en el *mercado*.

Aquí de nuevo y a riesgo de ser reiterativa me interesa señalar que, como ha podido verse en lo que va de esta breve exposición, en *La moral por acuerdo* se utiliza un vocabulario *económico* que considero un tanto ajeno al de los autores allí citados

(pp. 113-119). Creo que el lector bien puede apreciar que se produce un salto no del todo justificado entre un tópico del contractualismo clásico –la necesidad de salir de un supuesto estado de naturaleza mediante el recurso a un pacto social– y este nuevo tema súbitamente introducido por Gauthier: el *mercado*, sus *fallos*, y la propuesta de considerar a la moral como una superación de dichos fallos. Para ilustrar esto último, repárese en la forma en que nuestro autor describe cómo caemos en cuenta de la necesidad de la justicia. Esto ocurre según él, porque nos percatamos de las imperfecciones del *mercado* y, por esta vía, llegamos a ver en la justicia la solución que nos evitaría pagar los *costes* desproporcionados que nos impondrían *parásitos*, *rentistas* y *polizones*, quienes se eximirían de *pagar* lo que les correspondería en razón de los *beneficios* que obtendrían al *explotarnos*. Gauthier advierte entonces que, con Hume y Rawls, él también cree que son las insalvables condiciones humanas de la *escasez* en nuestro medio externo y del *autointerés* en nuestro carácter, aquellas que conforman las circunstancias que hacen necesaria la justicia. Como puede verse, en *La moral por acuerdo* los discursos del clásico escocés y del filósofo estadounidense son traducidos a un vocabulario *económico*⁶⁷, problema en el que de momento no insistiré más. Me detendré por ahora en la exposición de Gauthier acerca de las circunstancias de la justicia y del contrato social que pone fin a un supuesto estado de naturaleza.

Según el filósofo canadiense la racionalidad a la que acudimos cuando se pacta dicho contrato difiere de aquella otra de la que nos valemos cuando nos situamos en el estado de naturaleza descrito por Hobbes. En dicho estado, afirma Gauthier, cada agente elige una estrategia individual para resolver *su* problema, mientras que tratándose de una situación de interacción cooperativa los agentes –*todos*– tienen *el mismo* problema y ante él acuerdan actuar a partir de una estrategia común o acordada (*joint*) de cooperación (pp. 120 y 166).⁶⁸ Antes de ocuparnos de las características que debería tener esta estrategia conjunta, quisiera señalar que nuestro autor no parece tomar en consideración algo que ya se comentó en el anterior capítulo (en la referencia a Smith) y que bien podría ser aconsejable tener en cuenta en este intento de Gauthier por entrelazar justicia y mercado. Este último, como institución, creo que presupone que los

⁶⁷ Si bien habría que reconocer que en Hume la justicia, como virtud “artificial”, está unida a la defensa de la *propiedad*, de todas formas no es vista únicamente como una solución a problemas económicos. Cf. §§ 13-31 de la *Investigación sobre los principios de la moral*.

⁶⁸ Compárese en este punto: Axelrod (1984, pp. 73-87, 145-191; y 1997, pp. 40-120).

hombres ya han salido de un hipotético estado de naturaleza y, por tanto, que sus diversas prácticas son posibles gracias a las instituciones y reglas de juego que rigen dichas prácticas. Aunque, como también se ha mencionado, las imperfecciones del *mercado* pueden conllevar *injusticias*, sin embargo pienso que tal vez puede resultar poco acertado equiparar dichas imperfecciones a un supuesto *estado de naturaleza* como el que describen los padres del contractualismo moderno. De nuevo, en esto vemos otro elemento que tal vez admita ser juzgado como una inusitada traducción a términos económicos de un tema ya largamente trajinado en la historia de la filosofía política –el contrato social y como éste justifica un estado de derecho–, tema cuyos aspectos fundamentales tal vez se nos hagan borrosos si se lleva a cabo dicha traducción sin más, o sin que se elaboren al menos algunas mediaciones que la fundamenten y eviten que se caiga en anacronismos, sobre todo, tratándose del Hume.

No obstante esto último, debe señalarse también que en el desarrollo de la argumentación de *La moral por acuerdo* hay una suerte de giro. Éste se advierte en la importancia que adquiere el tema del contrato social *i.e.* del acuerdo a partir del cual se establecen las reglas de juego fundamentales para la interacción entre los ciudadanos. Si nuestro autor busca que tales reglas y el pacto que las origina sean *justos*, el giro argumentativo al que me refiero creo que se debe a la necesidad que tiene Gauthier de matizar su modelo de cooperación, hasta ahora centrado en el *mercado*, para que éste incluya de manera más evidente las virtudes morales del contrato social. De allí el anuncio que hace Gauthier a la altura del quinto capítulo: a partir de dicha parte de su texto hará un énfasis distinto que apunta a la necesidad de restricciones *morales* para una estructura de cooperación que las exige, en contraste con la estructura que supone el mercado perfectamente competitivo, el cual, como vimos, las hacía injustificadas a los ojos del autor. Este cambio implicará, según él, que va a tomar distancia de su insistencia inicial en el mercado y en los agentes económicos que interactúan en dicho escenario para centrarse, de ahora en adelante, en las características de un tipo de cooperación no coercitiva que gane la cualificación moral-racional necesaria para cumplir el papel del diseño básico de la interacción social y de sus principales instituciones. Ante esta advertencia del filósofo canadiense, por lo pronto simplemente me limito a dejar anotada una duda que bien podría asaltarle al lector y que ya se ha mencionado: acaso el nuevo énfasis que se anuncia y que se supone se fijará en lo moral

y en lo político, en lo que se refiere al problema de una cooperación justa y racional, ¿no haría innecesaria e, incluso desaconsejable, la insistencia inicial en el tema de lo económico y, más concretamente, en el modelo del mercado y de sus agentes egoístas? Dejo simplemente señalado este asunto que, como veremos en el siguiente capítulo, tendrá importantes consecuencias. Me concentraré por ahora en la cuestión de la necesidad y, por tanto, de la racionalidad de la cooperación y su relación con el tema de la justicia.

1.2 El principio del *minimax relative concession* MRC y la negociación en busca de un acuerdo *justo* y racional. La elección de una estrategia conjunta (*joint*) y, de nuevo, el problema de la motivación *moral*

El reto al que se propone responder Gauthier a partir del mencionado giro, estaría en establecer las condiciones para una elección racional que sea cooperativa, o para una elección cooperativa que sea racional y *moral*. Por tanto, que al ser vista como justificada por los actores en ella implicados, “se gane” sin violencia o cualificadamente la aceptación de éstos. Se trata de mostrar que la cooperación es una forma racional de interacción, o la respuesta racional al problema de la interacción humana. Para ello, por una parte, nuestro autor intentará hacernos ver qué permitiría que un sujeto enfrentado a los problemas que conlleva dicha interacción y, por ende, al que se le plantea la necesidad de cooperar con otros agentes, perciba como “racional” –i.e. conveniente para sus intereses– el entrar a *negociar* un *acuerdo* para *cooperar*. De otro lado, Gauthier también buscará establecer las características de aquello a lo que ya nos hemos referido como una estrategia conjunta o acordada (*joint*), en contraste con una de tipo meramente individual, como la que elegiría, según nuestro autor, un agente hobbesiano. Esto es importante pues las características propias de la primera forma de estrategia permitirían que ella se presente como racionalmente justificada a los ojos de los agentes que intentan ponerse de acuerdo para cooperar. Lo cual implica que, en virtud de tal estrategia conjunta, cada uno esté dispuesto a renunciar a una parte de los beneficios que podría obtener si no cooperara, es decir, si insistiera en maximizar sin restricciones su utilidad individual. Como ya se ha comentado en el capítulo anterior, esta suerte de renuncia requiere que cada quien prevea que el provecho que obtendría al cooperar y,

por ende, al ceder parte de lo que lograría si no lo hiciera, “pesará más” que aquello que conseguiría si no cooperara.

Por lo anterior, Gauthier se centra ahora en un tema que no había desarrollado en los primeros capítulos de su texto: la negociación (*bargaining*) en la que cada agente implicado estará dispuesto a participar con el fin de llegar a un acuerdo sobre los principios que deberían guiar la cooperación social. El autor advierte que a esta forma de negociación habría que atribuirle ciertas condiciones especiales. Toda negociación opera según la dinámica de iniciar pidiendo un máximo de beneficios para sí mismo a cambio de un mínimo de concesiones para los demás negociadores. Luego, poco a poco, se termina por pedir menos de lo que se exigía inicialmente y por ceder más de lo que se estaba dispuesto a dar al comenzar el proceso. Pues bien, para el autor, en el caso de la negociación que se lleva a cabo en busca del pacto social, este proceso ha de guiarse por un principio: el del “minimax en la concesión relativa” (*minimax relative concession*), al que en adelante llamaré MRC (pp. 142-145). El cumplimiento de este principio propuesto por el filósofo canadiense requiere que cada parte implicada en la negociación esté dispuesta a ceder teniendo en cuenta lo que cedan las demás, pero de tal manera que cada agente elija un resultado que le exija un mínimo de concesión que, a su vez, no sea mayor que el mínimo de concesión otorgado por los demás negociadores. A partir de aquí la conclusión importante que Gauthier trata de obtener es que el MRC constituye la clave de la conexión entre *racionalidad* y *justicia*, entre racionalidad y cooperación justa. Según el autor, es así como justicia y racionalidad coinciden: en el ideal de una interacción cooperativa que se haría posible gracias a unas restricciones imparciales a la conducta maximizadora de los agentes interesados en cooperar (pp. 150 y ss); restricciones basadas en el principio del MRC, el cual también sería la clave del acuerdo racional al que lleguen las partes luego de negociar.

La justicia sería, entonces, la disposición que cada parte tendría para conducirse según las restricciones acordadas, disposición que se hace necesaria si se quieren superar las consecuencias indeseables de las imperfecciones del *mercado*; es, pues, la respuesta racional a dichas imperfecciones. Como ya se mencionó, si se presentan estas últimas, no debería haber una zona libre de moral y, de este modo, la mano invisible que garantizaría la armonía de intereses en un mercado perfectamente competitivo, tiene que ser ahora reemplazada por esta otra mano “visible” –i.e. voluntariamente asumida– que

es la cooperación. Sólo ella sería capaz de producir la buscada armonía que podría lograrse en el contexto de los mercados con fallos, es decir, en ámbitos de interacción en los que se hace deseable que haya restricciones a la conducta maximizadora de los agentes.

Ante esta justificación de la moral como algo que surge a modo de solución a las imperfecciones de los mercados reales, creo que al lector bien pueden asaltarle las siguientes dudas: ¿qué incentivos tendrían quienes llevan la mejor parte en esos mercados, es decir, los rentistas, parásitos y demás explotadores, para restringirse en su conducta maximizadora y hacer caso de las restricciones morales, esto es, para cooperar con los demás agentes y suscribir con ellos, y sin engaños, un pacto social que sea realmente vinculante? Aquel que tiene asegurados sus beneficios en un mercado con fallos gracias, precisamente, a que no se restringe en su conducta maximizadora y explota su posición de dominio en el mercado, ¿qué incentivos puede tener para abandonar dicha posición? ¿Por qué renunciar a ella? Ante las dificultades que conllevarían estas cuestiones, bien puede uno preguntarse: ¿qué tan apropiado puede ser, entonces, seguir la estrategia de Gauthier de modelar la *moral* como el resultado de un acuerdo para paliar las consecuencias de las imperfecciones del *mercado*? Y, siguiendo este hilo conductor, ¿qué tan convincente puede ser, por tanto, explicar la elección de la conducta del cooperador (moral), como la respuesta de un maximizador-prudente o un maximizador en el largo plazo MR, respuesta que tal vez sólo sea estratégicamente explicable por parte de quien *no* se encuentre en la posición de dominio en el mercado?

En vista de lo anterior, me atrevo a proponer, en contra de la sugerencia de Gauthier, que tal vez aquello que explique, aunque no genética, sí conceptualmente, el que un sujeto decida ser moral, no tendría necesariamente por qué remitirnos a razones ancladas en cálculos de utilidad. Incluso creo que quien hiciera tales cálculos podría aparecer a los ojos de sus cooperadores como alguien de poco fiar en los terrenos morales. A lo mejor el ámbito del discurso en el que se sitúa la persona que intenta ofrecer las razones para abstenerse de explotar al prójimo, o las razones para hacerle reclamos a quien no se abstenga de hacerlo, *no* sea el mismo ámbito del discurso desde el cual se busca explicar (o invitar a, o disuadir de) una conducta maximizadora. Creo que si nos restringiéramos a este último tipo de discurso sería muy difícil denunciar o

fustigar con razones propiamente “morales” a quien no tiene incentivos para abandonar su posición de dominio en una interacción en la que haya favorecidos y desfavorecidos. Por ello pienso que resultaría un tanto complicado explicar, en general, cómo las razones que esgrimimos desde la moral deben poderse reducir, como intenta mostrar Gauthier, a razones que provienen del cálculo económico.⁶⁹

Un acuerdo de no-explotación mutua, o de mutua lealtad, bien podría ser el resultado de una interacción que no tendría por qué considerarse necesariamente “moral”. Por ejemplo, aquella que se daría a nivel de organizaciones delictivas.⁷⁰ Tal vez a Gauthier pueda traerle complicaciones para su propósito de hacer convincente su fundamentación de la moral y explicar la motivación que pudieran tener los agentes para ser morales, el hecho de que intente modelar el espacio en el que opera la moral, como el campo de interacción de unos negociadores que se ven obligados a pactar dadas las razones que harían pactar a quienes vean que pueden “ganar” más con un acuerdo que con el *statu quo*; y ello suponiendo que no surjan nuevos incentivos para traicionar el pacto. Este podría ser el caso de las razones que harían pactar a un grupo de mafiosos, o a un conjunto de grupos económicos disputándose un monopolio. Las razones que pueden aparecer como válidas a los ojos de este tipo de “cooperadores”, es posible que no sean consideradas como tales por parte de otro tipo de agentes: aquellos que no estén dispuestos a modelar sus interacciones al modo que parecen ser modeladas en la propuesta de Gauthier. Finalmente, creo que preocupa el que algo, o mucho, de lo que ocurre en esas interacciones entre agentes morales, y que puede ser importante a los ojos de quienes allí participan, se quede por fuera de la idea que el canadiense nos presenta de la moral, y de su explicación acerca de qué es lo que permite que haya sujetos interesados en tomar parte en la praxis moral. En este caso, la preocupación que puede a uno asaltarle pienso que tiene que ver con lo poco fiable o pertinente que pueda resultar su modelo, bien sea para dar cuenta de la moral como praxis, bien sea para dar razón de su base normativa.

1.3 Las razones para la justicia o para cumplir el pacto social; las respuestas de Hobbes y de Gauthier al “insensato”. Cómo sobrevivir en una sociedad de

⁶⁹ Si me permite hacer un uso laxo de la jerga wittgensteiniana, tal vez estaríamos ante dos juegos de lenguaje distintos e irreductibles entre sí.

⁷⁰ Véase: Catanzaro (1985); Gambetta (1993); Smith, D. (1975); Graebner Anderson, A. (1979).

MMII si se es un MR, y el problema del (no)carácter moral de los supervivientes

Sin duda, Gauthier advierte las dificultades que se le plantean al tratar de dar cuenta de las razones para la moral partiendo de una lógica maximizadora. De allí que de manera explícita anuncie que el aporte que él se propone ofrecer será justamente la demostración de que cualquier agente racional que se halle ante la disyuntiva de aceptar restricciones a su conducta maximizadora a cambio de los beneficios de la cooperación o, por el contrario, rechazar tales restricciones y por consiguiente correr los riesgos de la no-cooperación, necesariamente escogería la primera alternativa. Para ello el agente elegiría guiarse por el principio del MRC y participar en prácticas guiadas por dicho principio. Nuestro autor advierte que si él fallase en su intento de demostrar esta tesis, esto implicaría reconocer que la expresión “moral racional” carecería de sentido; que hablar de ella sería cosa vana o que la cooperación, aunque se base en un acuerdo racional, no sería requerida racionalmente de suyo, dado que no nos ayudaría a resolver el problema que plantean los fallos del mercado. En otros términos, el resultado sería reconocer que no puede señalarse una base racional de la moral, puesto que no sería factible ofrecer una justificación aceptable para las restricciones a la búsqueda de la utilidad individual (p. 158). Pero según Gauthier si, por el contrario, él demuestra que hay razones –ancladas en el autointerés– para ser justos y hacer honor a un acuerdo racional sobre las restricciones impuestas por la justicia, esa base racional para la moral sería indudable.

Para lograr hacer plausible esta demostración habría que atender a las críticas que provendrían precisamente de aquellas voces que más suelen asociarse a todo intento de dar cuenta de la justicia como nacida del autointerés, en especial, la voz de Hobbes. Este último, según el autor canadiense, advierte que justicia y razón no siempre o no necesariamente coinciden (pp. 160-161): los acuerdos, como arriba se mencionó, a veces se guardan, pero a veces se rompen atendiendo al llamado de la conveniencia individual. Luego, ¿cómo garantizar que quienes suscriban un pacto sean leales a éste? Gauthier cree que se podrían alegar en favor de la elección de ser cooperadores –y justos– las mismas razones aducidas por el autor del *Leviatán*. A saber, que aquellos que están dispuestos a traicionar acuerdos no son bien recibidos en la sociedad, la cual

indispensablemente funciona sobre la base de tales acuerdos.⁷¹ Y viceversa, una persona que se caracteriza por estar dispuesta a hacer honor a sus pactos dadas, por supuesto, unas condiciones de seguridad y paz, resulta ser elegible como compañera o cooperante por parte de los demás agentes; de manera que esa misma disposición a ser justa le ha de resultar de lo más beneficiosa (p. 162).

Estas razones son parte de la respuesta que intenta darle Hobbes al “insensato” (*foole*) para mostrarle que el ser justos “paga”.⁷² Pero, en opinión del autor canadiense, esta respuesta presenta el problema de que es fundamentalmente *política* en tanto que apela a mecanismos coercitivos que fuercen la tendencia a la justicia; concretamente, a la figura de un soberano en el que los agentes han delegado su derecho al uso de la violencia y que, por ende, cuenta con el poder para hacer cumplir los acuerdos o disuadir a quien tenga la tentación de no hacerlo. Esta instancia sería, entonces, en términos de Gauthier, ya no una *mano* —que operaría de manera *no coercitiva*— sino un *pie visible* (p. 163), puesto que tiene el poder para usar la fuerza con el fin de obligar —*coercitivamente*— a cada agente que, buscando su propio interés, también aporte a ese fin que todos buscan, y que no podría lograrse si no hubiese paz social. Para nuestro autor el poder que monopoliza el soberano de Hobbes lograría que la interacción social se estructure de manera análoga a como funcionaría aquella otra forma de interacción que se guíe por el principio del MRC propuesto por Gauthier. Esto es, haría que cada agente maximice su utilidad hasta un punto que sea compatible con la búsqueda que todos los demás hacen de su propio beneficio (p. 163).

No obstante el filósofo canadiense cree que ese “pie visible” hobbesiano resulta ser una solución muy costosa frente a los problemas causados por las imperfecciones del mercado, si se lo compara con la “mano visible” que aporta la moral tal y como la plantea Gauthier. De allí que éste afirme que la solución tendría que ser *moral* y no política,⁷³ a diferencia de la que, según él, aparece en el esquema de Hobbes. Esto implica que la salida ofrecida por Gauthier ha de ser pensada como voluntariamente

⁷¹ Cf. Hobbes, Op. cit, Cap. XV.

⁷² Ibid.

⁷³ En este punto (pp. 163 y ss.), Gauthier dice también separarse del énfasis que hace Rawls en lo *político* y su renuncia a tratar los temas de la justicia y del contrato social en términos *morales* (Rawls, 1996, pp. 11 y ss.). Como ya se ha mencionado, y se verá de manera más clara en lo que sigue, al lector acaso pueden asaltarle dudas sobre el carácter propiamente *moral* del pacto social que propone Gauthier, así como sobre sus críticas al autor del *Liberalismo Político*, y a los clásicos del contractualismo.

acogida por los agentes cooperadores, lo cual significa que no debe ser simplemente una imposición ejercida por una instancia externa a dichos agentes (p. 164). Habría que darle, entonces, otra suerte de respuesta al “insensato” a quien trata de convencer Hobbes; y el filósofo canadiense justamente se propone dar esa respuesta alternativa.

Dicha respuesta equivale a mostrarle a un actor racional por qué es mejor ser un maximizador restringido MR (*constrained maximizer*) que ser un maximizador irrestricto MI (*straightforward maximizer*), es decir, por qué resulta más conveniente basar las propias acciones en una estrategia conjunta (*joint*), renunciando a alguna estrategia individual que pudiera traer una mayor utilidad esperada y ello, dado el supuesto de que se presenten ciertas condiciones que hagan racional la opción por la estrategia conjunta. Parte de estas condiciones se refieren a las características que Gauthier atribuye a su maximizador MR. Se trata de un agente que puede percibir si sus congéneres están dispuestos a cooperar y que, de una manera “condicional”, hace uso de su propia tendencia a la cooperación (p. 169). De este modo, si se encontrara con un maximizador MI, entonces el MR no se animaría a cooperar con él, pues eso significaría el riesgo de ser explotado por el MI y, en unas circunstancias así, el MR haría mejor si optara por una estrategia individual en vez de una conjunta. Por contra, si un MR percibe que se encuentra ante otro MR como ella/él, en ese caso sí optaría por cooperar, esto es, por elegir de acuerdo a una estrategia conjunta. Este modo de actuar sería el que adoptaría el MR en situaciones de dilema del prisionero; de allí que este personaje pueda obtener, según Gauthier, unos beneficios que no serían accesibles para el MI, por más astuto y/o previsivo que pueda ser este último.

Ahora bien, puestos ante la pregunta acerca de quién de los dos, si el MR o el MI, a pesar de ser ambos racionales (cosa que el autor no pone en duda) sería, sin embargo, el que propiamente acierta al adoptar su estrategia de elección, Gauthier responde que la disposición a cooperar *depende* también de la disposición que *se perciba en los demás* agentes. Esta capacidad para “percibir” es fundamental, pues buena parte de la elección de la estrategia a seguir supone que pueda “verse” si los demás agentes están dispuestos a cooperar, si bien para nuestro filósofo este “ver” no puede significar que todos seamos absolutamente transparentes ante la mirada de los otros; lo cual sería poco realista. Más bien, lo que sí debe suponerse y se necesitaría como condición para elegir una estrategia de elección, ya sea la del MR, ya la del MI, es

que seamos más o menos “traslúcidos” los unos para los otros: ni completamente opacos, ni completamente transparentes. Ello les permite a los agentes racionales, tal y como los presenta Gauthier, el que puedan acceder a una solución *moral* a su problema de la cooperación y, por ende, el que no estén condenados a una salida meramente *política* como, en cambio, sí les sucede, según lo que afirma el autor canadiense, a los agentes completamente opacos que aparecen en el relato de Hobbes.

[. . .] supposing that persons are neither transparent nor opaque, so that their disposition to co-operate or not may be ascertained by others, not with certainty, but as more than mere guess-work. Opaque beings would be condemned to seek political solutions for those problems of natural interaction that could not be met by the market. But we shall show that for beings as translucent as we may reasonably consider ourselves to be moral solutions are rationally available (p. 174).⁷⁴

De manera que, supuesta esta condición de “translucidez”, los MMRR y los MMII a veces acertarán y a veces no, cuando intenten reconocerse entre sí. Los MMRR en ocasiones fallarán al no reconocer a otros MMRR con los que les conviene cooperar y, en ese caso, perderán las oportunidades ofrecidas por dicha cooperación; o por el contrario, acertarán al reconocerse unos a otros, obteniendo entonces los beneficios alcanzables en dichas circunstancias. En otras ocasiones podrían cometer el error de no reconocer a los MMII y por tanto correrán el riesgo de ser explotados por estos últimos. Gauthier trata de ofrecer un catálogo de las condiciones en las cuales un agente traslúcido (no transparente ni opaco) elige racional o correctamente qué disposición seguir, si la del MR o la del MI (pp. 174 y ss). Dichas condiciones dependerán, tanto de la disposición que se percibe en los otros agentes –también translúcidos–, como de las características de las circunstancias en las que se encuentre quien intenta hacer tal elección.

En este punto nuestro autor introduce otra de sus dicotómicas diferencias, según la cual habría que distinguir entre dos tipos de agentes. Por un lado, está el caso de una persona que tiene una tendencia a ser más dócil, a mostrar de manera *amplia* su aquiescencia, su tendencia a confiar en los otros, su voluntad de negociar con ellos y

⁷⁴ Como ya se ha dicho en un pie de página anterior, creo que a despecho de lo que aquí afirma Gauthier, en este aparte (y en lo que sigue) puede apreciarse que nuestro autor presenta más bien un cuadro *político* bastante similar al que él mismo le critica a Hobbes, antes que un modelo propiamente *moral* de la actitud cooperadora.

cumplir sus tratos o, simplemente, de ceder ante ellos (*broad compliance*). En contraste con esta actitud, hallamos la de otro tipo de agente, que tiende a mostrar de una manera más *restringida* y condicional esta tendencia asociada a la docilidad, la aquiescencia, la disposición a cumplir los acuerdos o ceder ante otros (*narrow compliance*; pp. 178 y ss). Una persona que muestra la primera actitud puede que se beneficie de situaciones en las que otro individuo, que manifiesta la tendencia opuesta, no pueda hacerlo. Pero la mayoría de las veces va a salir perjudicada a causa de esa disposición suya, por la que va a ser conocida entre los demás agentes, lo cual llevará a que muchos busquen la forma de explotarle. En cambio, un agente que tienda a ser dócil, negociador, complaciente o cumplidor únicamente de manera restringida o condicional (*narrowly*), según Gauthier, aunque igualmente se guíe por el principio del MRC, lo hará de tal forma que cooperará sólo cuando las situaciones en las que se encuentre prometan ser beneficiosas para todos los implicados y lo sean de manera justa; es decir, se comportará tal y como lo haría un astuto o razonable MR. En otros términos, un MR se caracteriza por aplicar el principio MRC y por mostrar una tendencia *restringida* o *condicional* a cooperar.

Lo razonable, pues, es adoptar la posición de un MR sólo cuando uno se encuentre rodeado de otros MMRR y en situaciones que favorezcan a este tipo de agentes, pero nunca si se está en un contexto que favorece a los MMII, los cuales serían los únicos que se benefician de la cooperación social en este tipo de circunstancias. La “virtud” o el “arte” de un MR inteligente –si bien Gauthier no utiliza estos términos– reside, entonces, en poder detectar la sinceridad de sus congéneres, logrando así acertar al identificarlos o como MMRR, o como MMII. Si un MR no tuviera esta habilidad perdería las oportunidades de beneficiarse de la interacción con otros MMRR, o caería en las trampas de los MMII. Si ocurre que uno es un MR y tiene la mala fortuna de vivir en una sociedad de MMII, la actitud razonable es la de *no* cooperar, la de buscar sólo la propia conservación y, por tanto, la de no cumplir pactos morales.

In a world of fools, it would not pay to be a constrained maximizer and to comply with one's agreements. In such circumstances it would not be rational to be moral (pp. 181-182).

Ya el lector habrá advertido los problemas que podrían señalársele a este, en mi opinión, perturbador retrato tanto del carácter moral –o, mejor, de la ausencia de éste–, como del tipo de comunidad moral que conformarían los agentes así descritos.⁷⁵ Creo que una pintura como la que así propone Gauthier puede resultar muy sugerente como recetario para la supervivencia de quien se siente, parodiando la frase evangélica, como “oveja en medio de lobos”, pero no lo sería tratándose de quien esté interesado, o bien en construir vínculos con otros agentes, o bien en modelar sus experiencias morales o las de aquellos que compartan dicho interés. Pienso que tampoco sería un cuadro ilustrativo para quienes no busquen ver a sus congéneres como lobos, ni quieran asumir el papel de ovejas bien pertrechadas con pieles de lobos o dispuestas a sobrevivir al coste que se requiera. Entre otras razones, porque las actitudes que se asuman ante el prójimo pueden determinar buena parte de su conducta hacia nosotros. Habría que recordar los estudios empíricos que muestran que quien parte de la base de que no puede confiar en los demás, termina produciendo en ellos la respuesta “espejo” o proporcional hacia su desconfianza⁷⁶. Quien se considera a sí mismo “uno de los buenos” y asume que probablemente los demás no lo son, acaba por encontrarse con que éstos perciben como algo riesgoso para ellos el portarse con él como lo harían “los buenos”. Al final parece que lo que en todo caso resulta difícil es atribuirle una identidad moral, un carácter (*ethos*), unas convicciones, un criterio, o algún tipo de identidad o de consistencia moral al “bueno” y astuto MR que trata de sobrevivir en un mundo que bien podría estar plagado de MMII dispuestos a explotarle. Ello, entre otros motivos, porque tampoco sería fácil saber si alguien así es “bueno” o “justo” o “moral” en algún sentido filosóficamente interesante; o si alguien así es propiamente un agente moral.

Creo que concebir la comunidad moral como un sistema de cooperación restringida o condicional entre cooperantes restringidos y condicionales puede resultar, o bien una pintura hiperrealista de una sociedad de egoístas incapaces de tejer lazos con sus congéneres, o bien una pesadilla. Acaso lo que molesta es que la pesadilla ilustra o retrata más de lo que desearíamos a nuestras actuales sociedades avanzadas. De allí que

⁷⁵ En este punto tal vez resultaría pertinente recordar, entre otros aportes, la crítica de B. Williams (1972) al utilitarismo.

⁷⁶ Véase, por ejemplo, Axelrod (1984, pp. 60, 118-121 y 136-139).

podríamos convenir en que, si la intención de Gauthier fuese pintarnos fidedigna o descarnadamente, a lo mejor y, como digo, a pesar nuestro, esta pintura suya diría mucho acerca de aquello que consideramos ilustrador, por más terrible que nos resulte, de nuestra sociedad. Pero si esta no es la intención del autor, tal y como él se ha encargado de reiterar una y otra vez a lo largo de su exposición, entonces a lo mejor lo que desconcierte es que termina por decirnos muy poco acerca de qué es aquello en lo consistiría ser racionales en la vida moral; o cómo tendríamos que pensar a los agentes morales que intenten ser racionales; o cómo habría que concebir, en general, a un agente moral; o qué tipo de experiencias nos suelen caracterizar en tanto que tales agentes. Tal vez todo ello se deba a que Gauthier a estas alturas de su texto todavía sigue enmarcando su pintura del agente racional-moral dentro de los estrechos límites que, sin embargo, nos había prometido ampliar: los del *homo oeconomicus* que sólo se vincula con otros agentes económicos y opacos mediante relaciones contractuales o meramente estratégicas. En todo lo cual, por lo demás, sigue sin verse lo propiamente “moral” del asunto ni, por tanto, su relevancia ni tampoco las ventajas que ello supondría si se lo comparara con: 1) Los clásicos juegos de cooperación expuestos por los teóricos de la decisión o, 2) Con aquello que Gauthier ve como análisis puramente “políticos” de la cooperación social justa, v.g. el de Hobbes.

2. Un primer paso para la moralización del *homo oeconomicus*: su elección paramétrica de una disposición a actuar y su participación en el acuerdo para el contrato social

Tal vez previendo las críticas que acaban de mencionarse, el autor intenta corregir su anterior retrato del MR advirtiéndole que, en franco contraste con los teóricos de la elección racional, él piensa que un agente que maximiza su utilidad lo hace a partir de la escogencia, no de una *acción* concreta, sino de una *disposición* a actuar por la que opta paramétrica o no estratégicamente (pp. 182 y ss.). De allí que Gauthier reconozca que la disposición que tiene una persona a decidir de una determinada manera, afectará las situaciones en las que espere encontrarse. Así, un MR normalmente será incluido, mientras que un MI normalmente será excluido de los acuerdos a los que se llegue en su sociedad. De modo que el primer tipo de agente, mas no el segundo, es quien se beneficiará de su disposición a actuar, no tanto en virtud de las decisiones específicas

que haga gracias a tal disposición, sino más bien en razón de las oportunidades que de manera mediata, en el mediano y largo plazo, se abren ante él gracias a que ha optado por una determinada línea de conducta en lo que a sus elecciones se refiere. Estos cambios en el discurso del filósofo canadiense, a mi entender, pueden hacerlo más aceptable o más cercano a la investigación moral, al contrario de lo que sucedía con la pintura que antes ha mostrado nuestro autor. Creo que ello se debe, en buena parte, a que ahora parece que, en virtud de tales cambios, Gauthier tiene un poco más en cuenta el tema del carácter moral. Sin embargo también pienso que resultan claramente contradictorios con su postura previa.

A partir de estas nuevas afirmaciones Gauthier intentará separarse de los trazos más duros de su anterior retrato de la oveja dispuesta a comportarse como lobo⁷⁷ y buscará, entonces, llenar de contenido moral el modelo de lo que aparecía como un mero *homo oeconomicus*. En primer término, dice que se separa de Hobbes en tanto que el autor canadiense piensa que los agentes pueden *reflexionar y cambiar su disposición a elegir* de una cierta manera. Así, pueden dejar de ser MMII para convertirse en MMRR; pueden preocuparse por los resultados que tenga en el largo plazo su conducta recurrente de elección, de modo tal que la capacidad que tengan los individuos para cambiar su disposición a decidir de una forma determinada, revela la naturaleza no sólo racional, sino también *moral* de dichos agentes (pp. 183-184). Con esto último creo que Gauthier, si bien parece estar respondiendo a las críticas que se mencionaron al final del anterior apartado, no obstante también pienso que puede estar abriendo de nuevo la brecha entre “lo moral” y “lo racional”; brecha que, recordemos, nuestro autor se proponía cerrar mediante su programa general de deducir la moralidad de la racionalidad maximizadora. Así mismo, en mi opinión, resurge el ya señalado problema que se presentaba al haber tomado como punto de partida un compromiso tan fuerte con la figura del maximizador autointeresado, para luego intentar salvar los costes que esta misma figura impone, si se quiere presentar una idea plausible y no tan desagradable del agente moral.

⁷⁷ Gauthier no usa las expresiones oveja/lobo, pero creo que bien pueden utilizarse para describir el cuadro que acabamos de analizar.

2.1 El difícil equilibrio entre tres críticas: contra al emotivismo de Parfit, contra los “neo-kantianos” y contra la idea de una racionalidad no-maximizadora

Ante las anteriores objeciones y como parte de su toma de distancia frente a su inicial retrato del *homo oeconomicus*, nuestro autor insiste en que el problema que tales objeciones plantean no puede ser solucionado por aquello que él llama posiciones “neo-kantianas”. Pues éstas, en su opinión, fracasan al no mostrar una conexión entre la moral –obviamente *no* asumida en términos “tradicionales” o pre-filosóficos– y la racionalidad, a la que Gauthier sigue entendiendo como exclusivamente maximizadora; o mejor, entre la moral y los intereses y/o preferencias de los agentes. *Por una parte*, Gauthier reconoce que, en cierto modo, hay razones que justificarían los temores de aquellos a quienes él se refiere como “neo-kantianos”, dado que si se enfatiza el aspecto *maximizador* de la *racionalidad* y se la desliga del aspecto *restrictivo* de la *moral*, podría llegar a pensarse que esta última se halla por completo escindida de los intereses de los agentes y, por tanto, que a éstos, al ser seres racionales y, por definición, maximizadores, *no* se les podría presentar una justificación racional –sino, a lo sumo, emotiva– de la moral. Sin embargo, *por otro lado*, nuestro autor acusa a los “neo-kantianos” –y en mi opinión no le faltaría razón al hacerlo– de presentar un desgarrador cuadro de la agencia humana, al no poder fundamentar lo moral de tal manera que se evite una escisión insalvable entre el aspecto restrictivo de la moralidad y los intereses o las auténticas motivaciones de los agentes.

If the disposition to make straightforwardly maximizing choices were wired into us, we could not constrain our actions in the way required for morality. Moral philosophers have rightly been unwilling to accept the received interpretation of the relation between practical rationality and utility-maximization, because they have recognized that it left no place for a rational constraint on directly utility-maximizing behaviour, and so no place for morality as ordinary understood. But they have then turned to a neo-kantian account of rationality which has led them to dismiss the idea that those considerations that constitute a person's reasons for acting must bear some particular relationship to the person. They have failed to relate our nature as moral beings to our everyday concern with the fulfilment of our individual preferences. But we have shown how morality issues from that concern. When we correctly understand how utility-maximization is identified with practical rationality, we see that morality is an essential part of maximization (p. 184).

En este punto, creo que el lector bien podría estar de acuerdo con las críticas de Gauthier, en tanto que acertaría al señalarle a la moralidad entendida en términos a los que el autor califica de “neo-kantianos”, el problema de que ésta sería algo meramente *externo* e impuesto sobre (o contra) el individuo y sus auténticas motivaciones. Sin embargo, creo que tampoco el filósofo canadiense resulta más convincente que aquellos a quienes critica en este sentido, dado que él tampoco logra mostrar el vínculo que haría de lo moral parte de dichas motivaciones *internas*⁷⁸ del agente, en caso de que éste sea concebido como teniendo una identidad o un carácter moral; identidad o carácter que, como hemos visto, sigue sin hacerse presente en el discurso de Gauthier. Por ello creo que éste aún no han respondido las mencionadas objeciones de quienes tendrían, a mi entender, buenos motivos para echar en falta lo propiamente *moral* de su figura del maximizador MR. Pienso que si nuestro autor quisiera escapar a las críticas que arriba se le señalan y, a pesar de ello también ser un *cognitivista* ético, tendría que asumir de algún modo una posición que si él insistiera en no denominar “kantiana”, por lo menos sí tendría que ser de cierta forma “rawlsiana” o “habermasiana”, en virtud de la cual incluya la razón maximizadora dentro de una noción más amplia o plural de racionalidad. Con ello, lo normativo de la moralidad no se vería rechazado por el autointerés de los agentes, pero tampoco ocurriría que las auténticas motivaciones de éstos fuesen negadas por dicha normatividad. De este modo, se mostraría a la moral como algo que obedece a esa forma menos miope, menos básica o menos restringida de racionalidad, en tanto que no-reducida a los esquemas meramente maximizadores. De lo contrario, creo que nuestro autor no tendría más remedio que suscribir una suerte de *emotivismo*. Sin embargo, Gauthier pretende separarse tan radicalmente de éste, como lo hace de aquellas tesis a las que se refiere como “neo-kantianas”.

Para justificar su rechazo a las posturas emotivistas, el filósofo canadiense argumenta en contra de una propuesta a la que identifica como un claro ejemplo de estas últimas: la de D. Parfit. Gauthier intenta mostrar por qué carece de validez la idea de Parfit de que los agentes humanos, aunque es cierto que en algunas ocasiones podemos elegir constreñirnos en nuestra conducta maximizadora, esto *no* es algo que hagamos

⁷⁸ En relación con este asunto, resulta de mucha utilidad la aclaradora crítica de B. Williams contra aquellos que insisten en que la obligatoriedad moral procede de lo que dicho autor denomina meras motivaciones o razones *externas*, *i.e.* escindidas de los compromisos asumidos *internamente* por el agente moral (sus razones “internas”). Véase Williams (1995, pp. 35-45). Este asunto será objeto de especial atención en el cuarto capítulo.

por razones que obedezcan a alguna forma de estrategia maximizadora, sino que, en esos casos, nuestras motivaciones o bien pueden ser irracionales, o bien pueden tener su origen en *otra* forma –más amplia o distinta– de racionalidad.⁷⁹

An objector might grasp that it may be rational to dispose oneself to constrained maximization, but deny that the choices one is then disposed to make are rational. The objector claims that we have merely exhibited another instance of the rationality or not behaving rationally . . . (p. 184).

Gauthier, pues, no aceptaría la tesis de Parfit de que, ya sea por mor de racionalidad –tal vez entendida como algo más amplio o distinto que la racionalidad maximizadora–, ya sea por motivos no-rationales, en algunas ocasiones los agentes humanos pueden verse obligados a actuar de forma “irracional” en el sentido de no ser maximizadores. Para el autor canadiense, ese tipo de conductas se deben, no a un fallo ni a una estrechez de la racionalidad específicamente *maximizadora*, sino a una debilidad o imperfección de la propia razón humana en cuanto tal facultad; una facultad que bien puede verse superada por los problemas que trae consigo la interacción humana en general (pp. 184-185). Por tanto, el caso del individuo que decide ser un MR y de este modo renuncia a ser un MI, no se debe explicar, afirma Gauthier, apelándose a *otra* supuesta forma –según él espuria– de racionalidad. Luego, no sería correcto afirmar que el MR presenta un fallo como maximizador, o que no le basta con serlo y por eso se ve obligado a acudir a esa otra (supuesta) racionalidad. Es más, el asunto tampoco debe verse según nuestro autor, ni como un fallo ni, mucho menos, como una falencia causada por alguna forma de debilidad o cortedad de las capacidades racionales del agente. Lo que sucede es que la razón humana, entendida como una facultad y no como la mera instanciación de ésta en algunos individuos, presenta algunas imperfecciones que le son propias en virtud de sus funciones y de su naturaleza; imperfecciones que se agravan dada la complejidad de los retos que suelen presentárseles a todos los agentes racionales.

En este punto, creo que el lector bien podría preguntar: si se ha rechazado la tesis de Parfit, entonces ¿qué explica que un maximizador autointeresado y no concernido por sus congéneres desista de ser un MI y, a cambio de ello, quiera ser un

⁷⁹ Cf. Parfit (1984).

MR y, por tanto, busque participar en interacciones cooperativas con otros agentes, aún a costa de renunciar a ciertas ganancias o ventajas competitivas frente a ellos? ¿Dónde confluyen racionalidad y moralidad, confluencia que explicaría que alguien pueda ser al mismo tiempo un maximizador racional y un agente moral? ¿Cómo es que Gauthier justificaría su idea de que solamente puede ser considerada como propiamente “racional” aquella forma de moralidad que caracteriza a este tipo de maximizador MR? Para responder a estas preguntas nuestro autor intenta completar su anterior respuesta al “insensato” del relato hobbesiano, respuesta con la que, recordemos, se proponía mostrar el nexo entre ser racionales y ser morales. Con este fin, el de apuntalar aun más su argumento, Gauthier elabora su propia adaptación del *proviso* de Locke.⁸⁰

2.2 El *proviso* lockeano, la “igualdad en la racionalidad”, y cómo adaptar –en clave maximizadora– el rawlsiano punto de vista “arquimédico” para el contrato social

Para el autor canadiense, el *proviso* debe ser entendido como la cláusula de cumplimiento que está presente, tanto en la negociación que llevaría al acuerdo representado en el contrato social, como en la aplicación efectiva de los principios de dicho contrato en el desarrollo mismo de las relaciones sociales (las cuales tendrían lugar a partir de la firma del pacto). Gauthier (p. 203) propone la siguiente reformulación de la cláusula de Locke a partir de la reformulación que, a su vez, hace Nozick y según la cual se prohibiría mejorar la propia situación mediante cualquier tipo de interacción que dañe o empeore la situación de los demás, a menos que ello sea necesario para no empeorar la propia situación.⁸¹ Así, en las diversas formas de intercambio social, si bien cada agente queda eximido de la obligación de *ayudar* a los otros o mejorar la situación de éstos –a menos que se hallen ante un gran peligro, y ella/él pueda socorrerles sin exponerse a un riesgo mayor–, sin embargo, le queda prohibido *empeorarles* su situación tomando ventaja de ellos.⁸²

⁸⁰ Cf. Locke, *Segundo Tratado*, Cap. V, §§ 27-33.

⁸¹ Cf. Nozick (1990, pp. 177, 179-182). Salvo mi traducción y mi paráfrasis.

⁸² Tanto en esta interpretación suya de la cláusula lockeana, como en otros aspectos sobre los que nos detendremos más adelante, Gauthier dice hallarse lejos de Rawls y, por el contrario, cerca de Nozick, para quien, como veremos en breve, estamos eximidos de un deber de *beneficencia*, quedando sólo obligados a cumplir con un deber de *no-maleficencia*. Cf. Rawls, (1999, § 19), y Nozick (1990, pp. 184-196).

We interpret the Lockean proviso so that it prohibits worsening the situation of another person, except to avoid worsening one's own through interaction with that person. Or, we may conveniently say, that proviso prohibits bettering one's situation through interaction that worsens the situation of another. This, we claim, expresses the underlying idea of not taking advantage (p. 205).

Según el autor canadiense, lo que explica la estabilidad del pacto social y el respeto a la cláusula lockeana por parte de los agentes es la falta de incentivos para traicionar el *proviso*, dado que éste, al igual que el principio del MRC, exige un mínimo de sacrificio por parte de los individuos, a cambio de obtener los beneficios que se alcanzarían al cooperar. Un agente que no perciba que la actitud cooperativa y de respeto al *proviso* es la de todos los demás participantes, no tendrá, pues, incentivos para ser un MR. Pero si, por el contrario, percibe que sus compañeros adoptan la actitud de los MMRR, entonces él también tendrá en ello los alicientes para participar en interacciones cooperativas, pues éstas prometen los mayores beneficios para todos los que se involucren en ellas. De este modo el deseo de ahorrarse costes o de trasladárselos a otros individuos, se verá rebasado por el de obtener aquel tipo de beneficios que sólo serían accesibles para quienes participen en relaciones cooperativas en un contexto de paz social. Son esos beneficios de la cooperación los que inclinan la balanza para que el agente esté, por tanto, dispuesto a renunciar a aquellas otras ventajas que podría obtener si insistiera en ser un MI.

Why should an individual seeking to maximize his utility not take advantage of his fellows? Why does his interest in market and co-operative arrangements, in itself an expression of his concern with utility-maximization, afford him a reason to accept the proviso as limiting his rights in market and cooperative interaction? [. . .] Why should rational utility-maximizers interact in a fully cooperative manner? More generally, why should rational individuals enter fully into society, the locus of both market and co-operative practices, within an enduring state of nature? Indeed, is not the state of nature the underlying reality of the human condition? [. . .] We must show that without limitations that exclude the taking of advantage, a rational individual would not dispose himself to cooperative compliance. [. . .] It is not and cannot be rational to be more compliant than his fellows . . . since broad compliance is not rational for everyone, it is not rational for anyone. [. . .] The rationality of disposing oneself to narrow compliance, and so to the acceptance of the proviso, follows from the advantageousness of society and the equal rationality of its members. [. . .] In social interaction, requiring voluntary constraints, the desire to avoid costly interaction can be balanced with the desire to take benefits and displace costs. The proviso forbidding the taking of advantage, represents the weakest constraint rationally acceptable to persons who would avoid costly interaction with others, and the strongest constraint rationally acceptable to persons who would be free to benefit themselves . . . (pp. 223-227).

En todo lo anterior es fundamental, según Gauthier, que cada individuo involucrado en la cooperación perciba que tanto ella/él, como todos los agentes con quienes se relaciona, pueden desplegar *en igual medida* sus potencias racionales, esto es, que todos sean “igualmente racionales” (p. 226). Haciendo uso de esta nueva expresión acuñada por él, nuestro autor propone, entonces, un principio adicional, que va de la mano de los incentivos para hacer honor a la cláusula lockeana: el principio de “igualdad en la racionalidad”. Según éste, si unos agentes fueran “más racionales” que otros, en el sentido de poderse permitir el ser *más* maximizadores tipo MI, la interacción con ellos sería percibida como riesgosa por parte de aquellos otros agentes que no puedan permitírselo y, por tanto, a los ojos de estos últimos, no sería racional tomar parte en dicha forma de interacción. Pero si, por el contrario, la percepción generalizada que tienen los agentes es que en sus interacciones todos ellos pueden ser “igualmente racionales”, es decir, que todos pueden comportarse como MMRR y hacerlo en la misma medida sin sufrir perjuicios, entonces lo racional para todos será cooperar y, por esta vía, gozar de los bienes que produzca la vida en sociedad. De esta manera, el nexo entre ser racional y ser moral, o ser maximizador y aceptar las restricciones que la moral imponga a la búsqueda del propio beneficio, queda asegurado a partir del cumplimiento generalizado –por parte de *todos* los involucrados y *en igual medida* para todos ellos– de la cláusula de Locke (en la adaptación que hace de ella Gauthier). Y dicho cumplimiento expresa una *igualdad en la racionalidad* de todos los agentes que así cooperan.⁸³

Esta tesis apunta al intento de explicar por qué un maximizador podría estar interesado en ser moral y, por tanto, estaría dispuesto a renunciar a parte de sus beneficios, de forma tal que esta suerte de renuncia no sólo *no* implique que el agente

⁸³ En el siguiente aparte veremos cómo esta “igualdad en la racionalidad” puede ser cuestionada por parte de quienes sospechen del cumplimiento de dicho principio a nivel de las sociedades reales-históricas; lo cual podría poner en entredicho el nexo que Gauthier cree haber demostrado entre las razones para ser *moral* y las razones para *buscar el propio beneficio*. En este orden, podría sospecharse de la confianza que parece tener el autor en que todo agente racional tiene abierta ante sí la posibilidad de *negarse* a entrar en una interacción que perciba como *riesgosa* para sus intereses. Pues esto estaría aún por probarse, sobre todo si se tiene en mente el caso –real, y no meramente especulativo– de aquellos agentes que, aun a pesar de que pueden estar concientes de este peligro, sencillamente *no tienen el poder* suficiente para *negarse* a tomar parte en este tipo de interacción. En contraste con la situación de este tipo de sujeto “des-ventajado”, volveremos a plantear una pregunta que ya se ha hecho arriba: ¿por qué razón un agente que goza de una posición de *ventaja* o de poder sobre otro(s) individuo(s), se abstendría de sacar provecho de dicho(a) poder/ventaja?

deje de ser un maximizador, como tal vez pretenden, según Gauthier, tanto Parfit como los “neo-kantianos”, sino que, por el contrario, precisamente obedezca a dicha condición de maximizador y al tipo de racionalidad que le define. Adicionalmente, la pintura de este agente que es moral *por mor de utilidad*, tiene como contexto una sociedad de maximizadores tipo MR que se perciben a sí mismos y a sus congéneres como tales, y de allí la estabilidad de la que gozaría, al decir de nuestro autor, el contrato social suscrito por tales agentes. ¿Cómo pensar, entonces, tanto a esta clase de *sociedad*, como a los *principios* que han de guiarla a partir de lo pactado por sus integrantes? Para responder a esta pregunta, Gauthier intentará construir su modelo del *agente* ideal y de la *sociedad* que dicho agente conformaría con otros individuos que compartan sus mismas características. Y lo hará complementando su teoría de la *justicia* (cuyos contornos básicos ya han sido expuestos arriba) con su demostración de que dicha teoría tiene un anclaje necesario en la teoría de la *elección racional*. Con lo cual, para el filósofo canadiense, se introduce una clara separación entre su propuesta y la de Rawls. De allí que, para elaborar su teoría, Gauthier haga una adaptación de la idea rawlsiana de la “posición original” pero utilizando, en clave maximizadora, el concepto de un “punto de vista arquimédico”, intentado con ello, como se verá, establecer un hiato importante frente al autor del *Liberalismo político*.⁸⁴

El lector recordará que, según el filósofo estadounidense, para garantizar la imparcialidad de la negociación y del acuerdo sobre los principios que van a guiar las instituciones sociales básicas, así como la justicia de tales principios, es fundamental que los individuos situados en la posición original desconozcan el lugar que van a ocupar en la sociedad. Sólo así se la podría diseñar a ésta como un sistema de cooperación realmente justo. Por esto se hace necesario que estos agentes también ignoren los talentos con los que contarán y las demás circunstancias que resulten determinantes para el futuro desarrollo de sus vidas dentro de la comunidad que así conformen. Si bien Gauthier no comparte buena parte de esta idea rawlsiana del “velo de ignorancia”, sí está de acuerdo con la necesidad de postular el “punto de vista arquimédico” asumible por un agente que puede y debe ser imparcial al momento de emitir un juicio cualificado sobre los principios de justicia que gobernarían la sociedad a la que esté dispuesto a pertenecer. Así mismo, el autor canadiense considera de la mayor

⁸⁴ Cf. Rawls (1996, Conferencia I, § 4; Conferencia VIII, § 4).

importancia esa imparcialidad que debe caracterizar al agente pensado de este modo. Sin embargo, no cree que para ello se le deban hacer las exigencias impuestas por el velo de ignorancia de Rawls. Simplemente bastaría con que conserve algunos de sus rasgos como tal agente, rasgos que, en opinión de Gauthier, son indispensables para que el sujeto pueda deliberar y decidir como *cualquier individuo racional maximizador* que esté *tomando una decisión* que *afectará sus intereses*. Esto no significa, para nuestro autor, que su agente arquimédico deje de ser imparcial en comparación con los agentes rawlsianos. Por el contrario, también se trataría de un sujeto que *no* se constituye en un individuo específico o que posea singularidades que le diferencien de las demás personas como, en cambio, sí es el caso tratándose de cualquier agente empírico.

Lo que busca Gauthier es que su agente arquimédico sea igual de imparcial que el de Rawls, para lo cual, según el canadiense, debe conservar sólo aquellas características formales que le harían un ser capaz de negociar estratégicamente con otros agentes que sean igual de maximizadores que ella/él, esto es, con individuos que también busquen lograr lo que más conviene a sus intereses. Por tanto, los agentes pensados por Gauthier intentan igualmente deliberar, negociar y elegir racionalmente en pro de tales intereses. En suma, para nuestro autor y a diferencia de Rawls, se trata de que el sujeto ideal que se sitúe en el punto de vista arquimédico pueda *maximizar su utilidad* aunque, del mismo modo que los agentes rawlsianos, carezca de aquellas singularidades o elementos diferenciadores que podrían impedirle ser el sujeto verdaderamente imparcial que intente elegir aquellos principios de justicia que gobernarían su sociedad; principios que el sujeto va a estar dispuesto a cumplir y exigirle cumplir a los demás.

Her concern is necessarily impartial because it is based on the formal features of individual rational agency without the biasing content of a particular and determinant set of individual characteristics (p. 233).

Gauthier, a diferencia de Rawls, y de acuerdo con Nozick, no cree que un agente que intente cumplir con el *proviso* en cualquier circunstancia, deba interpretar la cláusula de Locke en el sentido de un deber de beneficencia sino, simplemente, de no-maleficencia. De allí que, también en contraste con lo que dice Rawls con respecto a la posición original, Gauthier aplica esta idea a la noción de un agente situado en el *punto*

de vista arquimédico, de modo que, para el filósofo canadiense, dicho agente no tendría que estar bajo un velo de ignorancia *tan “tupido”* como el que Gauthier cree que Rawls propone para quien asuma dicho punto de vista imparcial (pp. 219-220).⁸⁵ Pues, según el primero, en dicha situación los agentes no tendrían por qué desconocer ni sus talentos ni sus ventajas, al tiempo que tampoco tendrían por qué abstenerse de usar esos talentos y ventajas en su propio beneficio. No tienen, así mismo, el deber de usarlos en favor de otros; simplemente basta con que no empeoren la situación de sus congéneres. No obstante estas características que le diferencian del sujeto rawlsiano, Gauthier cree que su agente sigue siendo ideal, abstracto y absolutamente imparcial aunque, como maximizador, por definición sigue siendo un individuo básicamente no-concernido por otros agentes: “. . . who takes no interest in one another’s interests” (p. 234).

Ante la inquietud que pudiera suscitar este retrato que hace Gauthier de su agente arquimédico, en el cual pienso que bien podría verse, o un cuadro contradictorio marcado por tensiones insolubles, o un retroceso a la figura poco moral del agente económico, creo que una posible respuesta que nuestro autor podría ensayar sería la apelación a una suerte de agente kantiano. Esto es, proponer algo parecido al *homo noumenico* que puede situarse en el punto de vista moral pudiendo guiar su voluntad por los principios de la razón y desatendiendo, por tanto, aquellas inclinaciones suyas que podrían parcializarlo hacia sus propios intereses o hacia los intereses de esos otros agentes por quienes sintiera alguna simpatía especial. Sin embargo, no es esta la respuesta de Gauthier. Por el contrario, nuestro autor enfatiza que su modelo de agente, como aquel que puede situarse en el punto de vista arquimédico, no es en absoluto el del agente kantiano (p. 234).⁸⁶ Entre otros motivos porque la sociedad, tal y como la ha de diseñar el agente descrito por el filósofo canadiense, es pensada por su agente ideal como un *instrumento* para ella/él; es decir, como el medio más importante para lograr sus fines. Según Gauthier, si toda sociedad moldea a sus miembros de tal manera que determina las capacidades que éstos puedan llegar a desarrollar, entonces un individuo situado en el punto de vista arquimédico, al tratar de establecer la estructura

⁸⁵ Cf. Rawls (1999, §§ 24 y 41). Aquí uso términos como “tupido”, que Gauthier no utiliza, pero creo que con ello no estoy haciendo violencia contra la interpretación que hace nuestro autor del filósofo estadounidense.

⁸⁶ En esta misma línea tampoco sería el sujeto rawlsiano, situado tras un “tupido” velo de ignorancia (Gauthier, pp. 245-257).

institucional básica de esa sociedad que está intentando diseñar, debe preocuparse por lo determinante que dicha estructura será para el desarrollo y despliegue de las capacidades de los miembros de esa sociedad, incluyéndolo a él mismo. De allí que su interés en el diseño de las futuras instituciones sociales sea básicamente *instrumental*: se trata de asegurar un contexto que favorezca el logro de *sus fines*.

Como ya se ha dicho, los agentes situados en este punto de vista arquimédico deben, según Gauthier, negociar y suscribir el contrato social que establece los principios de justicia que han de guiar la sociedad, pero partiendo para ello del modelo de negociación, decisión e interacción que se seguiría de la *teoría de la elección racional*. Así, desde dicho modelo tienen también que poderse derivar las normas de interacción que gobiernen la sociedad de este modo diseñada, esto es, su marco institucional. Este último, según nuestro autor, debe fundarse sobre el principio de “igualdad en la racionalidad” y debe asegurar la defensa de éste, es decir, debe permitir que todos los agentes puedan ser *igualmente racionales*, estableciendo unas normas de interacción social pensadas para agentes maximizadores, e *igualmente maximizadores*. De lo contrario no habría una interacción imparcial o en condiciones de simetría y, sin ella, ya no habría incentivos para una conducta moral por parte de todos los agentes vinculados por dichas normas. El aporte importante de este esquema que propone Gauthier reside, según él, en que logra responder a los interrogantes antes señalados, ya que puede explicar cómo es que moralidad y racionalidad maximizadora pueden ser consistentes entre sí.

If morality and rationality are in harmony, then these principles must be those found in our earlier analysis. Archimedean choice must select individual expected utility-maximization, constrained by the proviso and minimax relative concession principle. The impartial perspective of the ideal actor must cohere with the perspectives of rational individuals actually engaged in strategic choice (p. 234).

2.3 El agente maximizador ideal de Gauthier vs. los agentes kantiano, utilitarista y humeano

Gauthier intentará mostrar que su agente ideal, situado en el punto de vista arquimédico, e interesado en establecer principios de justicia que sean imparciales, coincide con el agente maximizador involucrado en un juego estratégico, tal y como se

lo ha de pensar a partir del modelo de la teoría de la elección racional. Por tanto, según nuestro autor, se trata de un agente identificable con *cualquiera* de los actores racionales que tienen intereses particulares, que buscan maximizar su utilidad y que no están concernidos por otros agentes. Para hacer más clara esta figura que, en mi opinión, y como se mencionó arriba, bien podría aparecer como ambigua y dicotómica, el filósofo canadiense intentará desvincularla de los modelos kantiano, utilitarista y humeano de agente moral-racional. Gauthier advierte que, si bien su agente comparte con dichos modelos la característica de ser ideal y relativamente abstracto, en el sentido de que en su definición no entrarían contenidos concretos que den cuenta de su identidad y circunstancias específicas, sin embargo, dentro de dicha definición también se incluirían algunos elementos que quedan por fuera de los mencionados modelos. Más exactamente: aquellos aspectos que vienen dados por el hecho de que se trata de un maximizador auto-interesado y no-concernido por otros agentes; un sujeto con intereses propios, los cuales busca satisfacer siguiendo el principio de utilidad, pero no al modo utilitarista, pues no se interesa por el bienestar general. Si bien está dispuesto a respetar principios morales y, en este sentido, a restringir su conducta maximizadora, al igual que lo harían los agentes kantiano y utilitarista, sin embargo, esta restricción no se haría de una manera tal que el agente de Gauthier tenga que sacrificar sus propios intereses.

[. . .] She is aware of her individuality, but not of its particular content . . . As an actor, the ideal person is a utility-maximizer, seeking her greatest benefit. Her utilities may be independent of the utilities of others; the Archimedean point is compatible with the supposition of mutual unconcern. What is ideal is in large part determined by what is absent rather by what is present; the ideal actor is of course rational and generally informed, but she seeks her greatest utility without being aware of the particular circumstances in which she acts, the particular capacities with which she acts, and the particular preferences for whose fulfilment she acts (p. 235).

En mi opinión y como arriba se anunció, se advierte una tensión en esta pintura que nuestro autor hace de su agente ideal. Pues *por una parte* se intenta vaciarlo de contenido a fin de que sea precisamente eso, un ideal, y con el objeto de que no pierda la *imparcialidad* necesaria para que pueda decidir cualificadamente los principios de justicia que deben guiar la sociedad. Pero *por otra parte* Gauthier busca que conserve las características de un *maximizador* de utilidad auto-interesado y no concernido por otros agentes, y que estratégicamente se involucra con ellos en un juego de cooperación.

Moralidad, por un lado, y racionalidad estratégica, por el otro; carácter ideal e imparcialidad de una parte y, de la otra, autointerés y conducta maximizadora; son los polos de esta tensión que, a mi parecer, no le facilita a Gauthier el trabajo de dibujar una figura convincente de su actor ideal. Esto se complica aun más si se tiene en cuenta que nuestro autor explícitamente afirma que su agente ideal, situado en el punto de vista arquimédico, imparcial y ocupado en definir principios de justicia igualmente imparciales, sin embargo, debe seguirse por el *dictum* humeano de que “la razón es una esclava de las pasiones” (p. 236). En este aspecto, el filósofo canadiense se confiesa como completamente opuesto a la idea kantiana de una razón pura práctica que es indiferente –e incluso hostil– a las pasiones, si bien el agente ideal de Gauthier no sería tampoco el “simpatizante ideal” (*ideal sympathizer*) o el actor empático que proponía Hume, puesto que se trata de un ser no-concernido por sus congéneres.

La respuesta de nuestro autor ante las inquietudes que pudiera suscitar su figura –repito, acaso dicotómica– del agente arquimédico es que, si bien es un sujeto ideal y, por tanto, no se identifica con ningún individuo concreto, no obstante se trata de un actor *racional* que elige; que hace una *elección* con la que busca *maximizar* su utilidad individual, tal y como lo haría *cualquier* actor que, para favorecer sus propios intereses, siguiera los principios de elección racional. Esto es lo que, según Gauthier, hace que su agente sea *universal e ideal* y, por tanto, lo que permite que *cualquiera* de nosotros se *identifique* con él.

The Archimedean point nullifies the biasing effects of individuality while retaining the idea, not merely of individual choice, but of choice by an individual concerned to advance his own interest (p. 237).

Se trata, al parecer, de un agente que busca maximizar su utilidad individual, pero que no es ningún individuo en particular, o no se identifica, insisto, con ninguna individualidad en particular; alguien que posee una identidad –y es importante, como veremos, que la posea–, pero sin que se sepa de qué identidad se trata; un sujeto que pone sus capacidades racionales al servicio de sus pasiones, pero que desconoce qué pasiones tiene y que, en todo caso, no se apasiona por otros agentes. Creo que aquí el lector bien podría preguntarse, ¿qué ventajas comparativas tendría esta figura frente a los agentes ideales utilitarista, kantiano y humeano? Frente al agente de Gauthier,

pienso que los utilitaristas podrían alegar que su agente modélico poseería una preocupación claramente “más moral”, dado que elige las acciones que prometen un máximo de beneficio posible para todos, o para la mayoría de los afectados, y no sólo para ella/él. El agente utilitarista no pone sus propios intereses ni los de sus seres queridos por encima de los intereses de la mayoría de los afectados por una decisión suya; amén de que con ello se vería a salvo del asalto de pasiones que podrían viciar de parcialidad sus elecciones. Además de esta última ventaja alegada por los utilitaristas, creo que los kantianos podrían atribuirle a su agente modélico un carácter claramente más imparcial, dado que éste no tomaría sus decisiones pensando en beneficiar a nadie, sino en acatar un principio moral al que considera como igualmente justo para todos, independientemente de las consecuencias empíricas que traiga la aplicación de dicho principio. Incluso, pienso que alguien podría alegar que el agente “simpatizante” o empático de Hume, al igual que los agentes kantiano y utilitarista, posee una preocupación propiamente “más moral” si se lo compara con el agente de Gauthier, pues en éste predomina el auto-interés, la conducta maximizadora y el no-concernimiento por otros, con lo cual se hace un tanto difícil verlo como un agente propiamente “moral”.

Por supuesto, nuestro autor podría responder que él también intenta curar a su agente de la parcialidad a la que pudieran empujarle los sentimientos, y que sería el peligro que se conjura en los agentes ideales de Kant y del utilitarismo –aunque acaso no en el de Hume–; con lo cual, la base de la moral, tal y como la plantea Gauthier, sería *racional*, como proponen los primeros, en franco contraste con las ideas pre-filosóficas o populares de moralidad. Sin embargo, también creo que si ganara esta base racional, de todos modos la figura del agente que propone el filósofo canadiense no ganaría el cariz moral de los tres modelos criticados por él. Incluso, si nos atrevemos a ir más allá, opino que un lector algo quisquilloso podría acusar a Gauthier de que también su agente conserva una pasión que, a diferencia de los sentimientos morales del agente empático de Hume, sigue sin poder ganar sustancia moral, o tal vez sea la causa, paradójicamente, de que la pierda por completo: la pasión por sí mismo.

Esta pasión, si bien nuestro autor se negaría a considerarla como tal pasión, sigue presente incluso en la crítica que le hace a un utilitarista contemporáneo como Harsanyi, a quien Gauthier descalifica porque, según él, dicho autor concibe la elección

correcta como aquella con la que se buscaría maximizar una media de bienestar social (pp. 238-245).⁸⁷ En ello, el filósofo canadiense advierte del problema que supone el hecho de que el agente utilitarista de Harsanyi no esté preocupado por su *identidad*, mientras que el autor de *Morals by Agreement* cree que su agente sí debería estarlo, dado que sería aquel con el que *cualquier* sujeto real podría *identificarse* como un ideal de su propio yo (*self*).⁸⁸ Pues Gauthier afirma que en la figura de su agente ideal se halla el elemento que permite que se produzca esta suerte de *identificación* por parte de cada agente *particular*: la idea de la *elección* que haría un agente *racional*, preocupado por su *propio interés*: “What connects each person with the ideal choice is the idea of the choice of a rational actor” (p. 255). Sin embargo, no creo que sea fácil presentar esta idea de la elección hecha por un agente maximizador ideal como la clave de aquello que haría que *cualquier* sujeto *racional* se *identifique* con el agente modélico de Gauthier, por lo menos no en lo que se refiere al aspecto propiamente *moral* de la elección que llevaría a cabo quien se sitúe en el punto de vista arquimédico. Pienso que esta dificultad puede aparecer, a los ojos de un lector interesado en hallar la base racional de la moral, como otro de los artificios de Gauthier y, en este caso, un artificio orientado a lograr que el punto de vista imparcial-moral no dependa de una base de carácter “kantiano”, tal como aquella a la que acude Rawls,⁸⁹ sino que dependa del *dictum* de la maximización de utilidad, tal y como se propone desde la teoría de la elección racional.

Como ya se ha dicho, nuestro autor insiste en que esta última proporciona el modelo de su agente ideal y ello sin que éste pierda su carácter imparcial. Gauthier cree que con esto su propuesta presenta una ventaja considerable frente al modelo de Rawls, a quien el filósofo canadiense acusa de negar la importancia que debe dársele a la idea de la elección racional –en clave estratégica–, y de no poder conectar, por ello mismo, a su agente ideal con lo que cada actor individual estaría dispuesto a elegir si se situara en el punto de vista arquimédico (pp. 235-238; 244-245 y 255).⁹⁰ Gauthier afirma que

⁸⁷ Cf. Harsanyi, 1982 y 1985.

⁸⁸ Compárese esta noción de “identidad”, por ejemplo, con la que propone Christine Korsgaard (1996, *Lecture 3, The Authority of Reflection*, pp. 90-130) mediante su idea de “identidad práctica” (*practical identity*). En otra línea distinta, frente a la noción de “identidad” que aparece en Gauthier, cabría recordar la crítica de Sen y Williams al utilitarismo, en lo que se refiere a un olvido, por parte de éste, de la identidad, individualidad y separabilidad de cada agente (1982, pp. 1-38).

⁸⁹ Véase Rawls (1999, § 40: “La interpretación kantiana de la justicia como imparcialidad”).

⁹⁰ Cf. Rawls (1999, § 41).

en su propuesta, en cambio, sí puede apreciarse claramente dicha conexión, puesto que las características formales que presentaría su agente ideal pensado como un maximizador, son características comunes a las de cualquier actor racional real y, por esto, *cualquier agente empírico puede identificarse con su agente ideal*, lo cual es posible, repito, sin que este último pierda en absoluto su carácter imparcial. Si se me permite usar el vocabulario del propio Rawls, creo que en este punto Gauthier daría la impresión de querer fundir en una misma forma de racionalidad aquello que el filósofo estadounidense llama “lo racional” y “lo razonable” de los agentes ideales que intentan elegir los principios de justicia. Para lo cual, en el esquema de Gauthier es necesario que esa forma básica de racionalidad sea la *maximizadora* y que el velo de ignorancia “tupido” que Rawls propone, como arriba dije, sea reemplazado, por el autor canadiense, por un velo más “fino”; si bien sigue sin poder establecerse con claridad *qué tan fino* deba ser, o cómo convencernos de que realmente *hay un velo*, dado que podría concluirse que acaso no haya ninguno. Si se le concede a Gauthier que sí lo hay, sólo que se trata de un velo menos tupido, con esto último nuestro autor tal vez logre darle más juego a eso que Rawls llama “lo racional”⁹¹, lo cual, dicho en términos del autor canadiense, estaría representado por el peso de las preferencias individuales y de la maximización de la utilidad individual, aunque Gauthier insista en que esto no llegaría al punto de cargar de parcialidad a la figura de su agente ideal.

3. Las exigencias normativas del *proviso*, y las tensiones entre lo aplicable y lo ideal del modelo de Gauthier

Todos estos malabarismos hechos por nuestro autor para presentar como consistentes los diversos –y tal vez contradictorios– aspectos de su agente ideal, se intentan conciliar y resumir en las dos direcciones hacia las que apunta, según él, la elección hecha por dicho agente. Por *una parte*, éste se mueve en una dirección *maximizadora* en virtud de la cual sólo entrará en interacciones de las cuales espere obtener *beneficios*; o, mejor, en las que todos los partícipes resulten mutuamente beneficiados y nuestro agente, por tanto, no quede en desventaja frente a ninguno de

⁹¹ Como ya advertí en un pie de página anterior, Gauthier no habla en estos términos que me permito tomar del propio Rawls. Ahora bien, en este punto específico, creo que el filósofo canadiense no estaría de acuerdo con mi interpretación de su lectura de Rawls, interpretación que, sin embargo, confío que el lector encuentre justificada. Cf. Rawls (1996, Conferencia II, § 1).

ellos. Por *otra parte*, el sujeto también actúa en otra dirección que le lleva a elegir una “libertad básica” (*basic freedom*) para sí mismo y para los demás agentes; una *libertad* que les permita a todos interactuar de una manera tal que cada uno pueda obtener la satisfacción de sus deseos y necesidades. Así, las interacciones que establezca cada agente con sus congéneres, se orientarían a servir a sus *propios intereses* y no solamente al beneficio de los demás. De este modo se producen relaciones sociales no forzadas o *no coaccionadas* en las que la satisfacción obtenida por cada agente particular es, al mismo tiempo, compatible con la libertad de todos aquellos con quienes interactúe.

If the ideal person chooses to benefit herself, then she must choose freedom for all. In choosing among principles of interaction and the social structures that embody them, then, the ideal actor ensures first that each person begins with a fundamental freedom to fulfil his preferences or advance his interests, and second, that social institutions and practices are intended to afford mutual benefit. These two requirements –liberty of action and mutuality of benefit– are linked. [. . .] The ideal actor thus chooses to be free so that she may choose only genuinely beneficial interaction, and being free, she of course then chooses to interact only in beneficial ways. [. . .] In choosing freedom of action and mutual benefit through interaction, the ideal actor imposes only very weak and hardly controversial limitations on the range of acceptable social structures . . . (pp. 257-258).

Libertad y beneficio mutuo son, entonces, las dos exigencias normativas del *proviso* y, por tanto, deben ser protegidas por las instituciones sociales que diseñen los agentes ideales que deliberen situados en el punto de vista arquimédico.⁹²

3.1 El aspecto “aplicable-realista” de su modelo, según Gauthier

Gauthier intenta mostrar que la elección que, del modo descrito, hacen dichos agentes ideales sería, de alguna forma, parecida a la elección de los agentes hobbesianos. Recordemos que, según nuestro autor, éstos optan por la paz y por la cooperación social dado que en ello ven oportunidades de beneficio que les serían inaccesibles si permanecen en estado de guerra, al tiempo que ven los enormes costes que dicho estado les significa. Sin embargo, Gauthier intenta mostrar que sus agentes ideales, frente a los agentes hobbesianos, presentan la ventaja de un *plus* moral que viene dado por el hecho de que los primeros se caracterizan por una *permanente*

⁹² De nuevo dejo de lado la cuestión de si esta apuesta por la libertad y el mutuo beneficio es tan “débil” –en el sentido de neutralidad valorativa– como Gauthier afirma.

voluntad de cooperación, lo que va más allá de las relaciones contractuales que conformarían el ámbito al cual un lector crítico podría intentar reducir la relevancia o aplicabilidad del modelo de agente que Gauthier presenta. Esta nueva advertencia de nuestro autor se debe a que hasta este momento, en lo que va de su texto (p. 259), ha hecho un énfasis muy marcado en aquellos aspectos de su modelo de agente como un maximizador autointeresado y no-concernido por sus congéneres. Para suavizar dicho énfasis, Gauthier trata ahora de introducir la novedad de que su agente ideal tiene además, y en contraste con los agentes hobbesianos, una *permanente voluntad de interrelación* con sus semejantes, con la que supera los fríos cálculos de costes-beneficios que le llevaron a pactar con ellos y que, consecuentemente, le hace interesarse por cualquier forma de interacción que prometa iguales ventajas para todos lo implicados:

She [the ideal actor] exhibits a continuing willingness to engage in social interaction pursuing mutual advantage whenever and wherever it may be found (p. 259).

Esta permanente voluntad de cooperación hará que a la hora de escoger, tanto los principios que guíen, en general, la cooperación social, como aquellos otros que sean más específicos y que determinen los nuevos ámbitos en los que dicha cooperación vaya surgiendo con el tiempo, el agente ideal siempre elegirá tales principios a partir del *proviso*. Pues, según Gauthier, éste no sólo se constituye en una *restricción* que impone los mismos *límites* a todos los sujetos involucrados en los diversos modos de interacción social, sino que también es un *medio* igual de *útil* para cualquiera de ellos y, por tanto, es el instrumento que necesitan para la obtención de los beneficios que les traigan dichos modos de interacción (pp. 259-269). De esta manera, la cláusula de Locke (en la versión que de ella hace Gauthier) no sólo es condición inicial de cualquier forma justa de cooperación, sino que también aporta el marco general dentro del cual se deben diseñar las instituciones sociales y se han de desarrollar las prácticas asociadas a éstas; instituciones y prácticas que tendrían, por tanto, que propiciar las nuevas formas particulares de transacción que vayan surgiendo con la evolución social (p. 260).

Creo que un lector bien podría preguntarse qué tan realista resulta asignarle este, al parecer, tan extenso alcance al *proviso* lockeano, sobre todo si tenemos en cuenta las fronteras culturales y societales, y puestos ante la importancia que Gauthier dice

concederle a la *identidad* de los agentes. Ante esto, nuestro autor responde que su agente ideal no haría, en principio, una diferencia entre un “nosotros” y un “ustedes” (o un “ellos”) y que, por ende, no aplicaría las normas derivadas del *proviso* solamente dentro *su grupo*, sino que las haría extensivas para todo agente racional con el que pueda interrelacionarse. El actor ideal querría la paz y la cooperación mutuamente benéfica en cualquier contexto en el que se encuentre, de manera que no sería ni xenófobo, ni clasista, ni racista, ni sexista. Incluso Gauthier encuentra que este agente y los principios de justicia que elige, pueden salir airoso frente las críticas de aquellos marxistas y/o feministas que insistan en señalar que en el mundo real que conocemos, con seres humanos reales, no ideales, la constante *no* ha sido propiamente esta suerte de “reino de Dios en este mundo” (por supuesto, Gauthier no usa estas palabras); y que, por el contrario, como señalan dichos críticos, lo usual ha sido que unos individuos salgan favorecidos mientras que los otros son explotados o no reconocidos en sus interacciones sociales (pp. 260 y ss.).

La razón de que esta objeción no afecte, según el autor canadiense, a su modelo de agente ni a la plausibilidad de los principios de cooperación social que dicho agente elegiría, reside, precisamente, en que se trata de un *ideal* con carácter *normativo* y, por ello mismo, *no* puede resultar afectado por las circunstancias *históricas-concretas* que señalan marxistas y feministas (pp. 261-265).⁹³ Como arriba se comentó, nuestro autor propone su principio de “igualdad en la racionalidad” como parte indispensable de los incentivos para que un agente se decida a ser un MR, renunciando a los posibles beneficios de ser un MI pero obteniendo, a cambio, las ventajas de participar en la cooperación social. Con ello, recordemos, según Gauthier quedan asegurados los nexos entre maximización y moralidad, entre ser racionales y ser morales, salvándose así la tesis central de *La moral por acuerdo*. Sin embargo, dichos nexos son seriamente puestos en cuestión por las mencionadas objeciones que nuestro autor atribuye a críticos de corte marxista y feminista, quienes señalarían que resulta ingenuo (y hasta ideológico) sostener el principio de una *igualdad en la racionalidad* por parte de los agentes sociales, si se tienen en cuenta los hechos históricamente rastreables, incluso tratándose de nuestras sociedades occidentales contemporáneas. Pues a la luz de estos hechos, lo que se hace más evidente es que la posibilidad de desplegar sus capacidades

⁹³ Gauthier no especifica a qué autoras o autores se refiere cuando critica al marxismo y al feminismo.

racionales y de participar únicamente en aquel tipo de interacciones que les resulten beneficiosas, se muestra como un privilegio del que no todos los agentes sociales han podido o pueden gozar por igual. Así, según esta crítica, la igualdad que, al decir de Gauthier, garantizaría los vínculos entre moralidad y racionalidad maximizadora, estaría lejos de ser la regla general y, por tanto, podría acaso pensarse que la tesis de nuestro autor descansaría sobre una quimera (pp. 263, 317).⁹⁴

Frente a esta crítica, la respuesta de Gauthier se mueve en dos direcciones que, a mi entender, no parecen ser consistentes. Por *una parte*, y como ya se ha mencionado, nuestro autor afirma que las situaciones históricas de desigualdad pueden presentarse al nivel, no del modelo *ideal* de cooperación racional, sino al nivel de los mercados y de las sociedades *reales*. El filósofo canadiense reconoce que en estas últimas puede ocurrir que se rompa el nexo entre moralidad y racionalidad cuando es el caso que algunos agentes tienen la posibilidad de desplegar su racionalidad y otros no; de modo que los primeros dictarían los términos del pacto social en contra de los intereses de los segundos. Sin embargo, frente a las críticas que señalan estas situaciones de injusticia real, el carácter ideal de su modelo seguiría conservando, a los ojos de Gauthier, su fuerza normativa quedando, por tanto, blindado frente a tales críticas, dado precisamente que se trata de un ideal al que se apela para denunciar y tratar de reparar esos hechos históricos señalados por marxistas y/o feministas. Pero *por otra parte*, nuestro autor afirma que habría que tener en cuenta la *naturaleza específica de algunos* de los *factores empírico-históricos* que producen las situaciones de desigualdad en la racionalidad de los actores sociales y, por tanto, la ruptura que a nivel de los hechos se produce entre racionalidad y moralidad.

En este sentido según Gauthier, puede haber factores *ideológicos*, representados por ciertas creencias que inciden en el hecho de que en algunas sociedades se presente

⁹⁴ En el siguiente capítulo se verá con mayor detenimiento esta crítica, en una versión bastante parecida: aquella que plantean los sofistas, según lo que narra Platón en la *República*. En dicho texto, analizado por Gauthier, el filósofo griego pone en boca de Trasímaco y Glaucón la tesis a la que intenta responder Sócrates, y que, como veremos, Gauthier también necesita desautorizar. A saber: que la justicia es sólo un mal necesario, o un mal menor al que apelan los débiles cuando no pueden rebelarse del todo contra los poderosos, o una artimaña a la que recurren estos últimos cuando no pueden dominar eficientemente a los primeros mediante el mero uso de la fuerza y, por tanto, acuden a la manipulación ideológica. Si a partir de allí se intenta concluir que sería ilusorio sostener un auténtico nexo entre ser racionales y ser morales, y si para Gauthier, por el contrario, es fundamental demostrar tal vínculo, entonces nuestro autor tendría que salir airoso frente a este tipo de críticas, por ahora puestas por él en boca de marxistas y feministas y luego, como veremos, recordando los viejos temas sofistas y hobbesianos.

una marcada diferencia de estatus entre algunos de sus miembros, por ejemplo, entre hombres y mujeres; y esta diferencia haría que a los primeros se les concedan oportunidades y roles que a las segundas se les niegan (pp. 230-232, 351-353). Las enormes dificultades que habría para cambiar estas creencias, sobre todo si son de mucha influencia en las sociedades en cuestión, y vienen precedidas de una larga tradición, harían que las mujeres no tuvieran más remedio que cooperar y cumplir con su parte de un anómalo “pacto” social, esto es, se vieran forzadas a ello dado que podrían ser objeto de castigo si se negaran a hacerlo. Obviamente en este caso su aquiescencia ante las normas no vendría atada, en sentido estricto, a consideraciones sobre los beneficios que obtendrían de la cooperación, esto es, no estaría propiamente ligada a la racionalidad.

El otro tipo de factor que tendería a romper el nexo entre esta última y la moralidad, vendría dado por las *diferencias existentes entre sociedades* que gozan de un gran desarrollo tecnológico y aquellas otras que carecen de él: “Technology is power. Those with a more advanced technology are frequently in a position to dictate the terms of interaction to their fellows” (p. 231). Para ilustrar esto último, Gauthier propone como ejemplo las diferencias, en cuanto a desarrollo tecnológico, que caracterizaron a los conquistadores españoles frente a los pueblos indígenas americanos (pp. 231, 290-294). Creo que resulta un tanto difícil no acusar a nuestro autor de hacer un uso ideológico de este ejemplo con el fin de hacer otra de sus apologías de la sociedad occidental como moralmente superior a otras sociedades. Por momentos parece que Gauthier sugiere que el criterio para determinar la mayor o menor racionalidad de un agente, en este caso, un agente colectivo (los conquistadores), viene dado por su *desarrollo tecnológico*. Así, racionalidad se identificaría con tecnología, y la capacidad de hacerse con una posición de dominio (militar o político) sería el indicador de una *mayor racionalidad* por parte de quien domina. Claramente esto supone que la moral, como restricciones al autointerés, sería una carga impuesta por quien domina a quien es objeto de dominación –v.g. las mujeres, los indígenas– y, con ello, lo que sucedería es que la base racional de la moral quedaría desautorizada; resultado que, como puede verse, no favorece en absoluto la tesis de Gauthier.

Sin embargo, nuestro autor se apresura a contestar que la *igualdad en cuanto a la racionalidad* de los agentes, que es un supuesto para que dicha base racional quede

asegurada, sigue siendo sostenible *por más que haya situaciones históricas que parezcan contradecirlo*. Pues esas situaciones históricas son *cosa pasajera*, que se superará a medida que las falsas creencias vayan cayendo en el descrédito y las diversas sociedades imiten cada vez mejor al modelo occidental de desarrollo tecnológico. A medida que avance esta suerte de proceso global de ilustración, cuyo núcleo es la tecnología como criterio tanto de racionalidad, como de libertad y de moralidad, entonces cada vez se hará más claro que todos podemos ser considerados como “igualmente racionales” y, con ello según Gauthier, no se está apelando a la fe en alguna quimera.⁹⁵ El destino de la humanidad, esta suerte de “fin de la historia”⁹⁶ estaría, pues, en imitar exitosamente el progresivo desarrollo tecnológico de Occidente y, con él, el progresivo desmonte de las creencias irracionales. Según nuestro autor, es sólo cuestión de tiempo que este proceso se haga extensivo a todas las sociedades y, así, éstas poco a poco irán asimilándose a la moderna sociedad occidental, la cual se ha caracterizado por el despliegue de un progreso tecnológico que resulta indesligable de su progreso moral. Es por ello que, a los ojos de Gauthier, no debe concedérsele mayor importancia a las críticas que enfatizan las desigualdades fácticas entre aquellos grupos o entre aquellas sociedades que están en posición de favorecer sus intereses –pudiendo, por tanto, permitirse el ser más maximizadoras *i.e.* “más racionales”– y aquellos otros grupos o sociedades que, por causas empíricamente superables, no pueden, por el momento, permitírselo; pues todos, en lo esencial, y así ha de confirmarlo la historia, somos igualmente racionales.

Pienso que a los ojos del lector no sería necesario que entre a comentar esta nueva apología de Gauthier. Por lo pronto, me limito a llamar la atención sobre el hecho de que se equipare desarrollo *racional* a desarrollo *tecnológico*, y posición de *dominio* a *mayor racionalidad*, pues creo que resulta ser muy ilustrativa del tipo de *racionalidad* y del tipo de *agente* modélico que se privilegia en la propuesta de Gauthier. No puedo

⁹⁵ “We suppose that the unequal rationality brought about by technological differences between societies is accidental. It does not reflect underlying differences in the capacities of persons constituting those societies. Freed from false views of the world and the practices and institutions to which such false views give rise, human beings tend toward technological equality. They tend, then, toward a state of affairs in which the proviso, the principle of minimax relative concession, and the disposition to narrow compliance, come fully into play in linking the requirements of reason, to the demands of impartiality of morals” (p. 231).

⁹⁶ Gauthier no lo dice usando esta expresión, pero suscribe, como puede ver el lector, una tesis cercana a la de Fukuyama. Cf. Fukuyama (1992).

detenerme, repito, a comentar los aspectos morales y políticos que sin duda habrán llamado la atención del lector –v.g. el cariz entre inocente e ideológico de las palabras de Gauthier, o su insensibilidad ante la cuestión de la desigualdad, lo cual explica, por ejemplo, buena parte de sus críticas a Rawls y sus simpatías por Nozick–. Pero sí quisiera dejar anotada esta cuestión, en cuanto creo que refleja de manera muy notoria y muestra los rasgos menos alentadores –como sucedía con la metáfora de Robinson–, del ideal de un agente *maximizador* y *no-concernido* por otros. Por ende, esto haría que se volviera a plantear el ya mencionado problema de fundamentar lo moral en términos *económicos* y, ahora, adicionalmente, como un asunto relacionado con el *desarrollo tecnológico* y el *poder* (político y militar), pasándose, de nuevo, por alto el escollo que supone el postular el *statu quo* de ciertas sociedades como moralmente ejemplar.

3.2 El mercado como el resultado tanto ideal-conceptual, como histórico, de la aplicación del *proviso*

Por lo anterior, creo que a Gauthier se le haría más fácil defender la plausibilidad de su modelo si le asignara el alcance que tienen todos los ideales y, por tanto, si hiciera énfasis en su aspecto crítico y *normativo* –tal y como sí lo hace, por ejemplo, Rawls, a quien tanto desautoriza el filósofo canadiense–. En otros términos, pienso que, de este modo, nuestro autor tendría más éxito en su empeño por blindar su propuesta frente a las mencionadas críticas marxistas y feministas, que le echan en cara los *hechos* como prueba en contra de su modelo. Pero este posible éxito tal vez se lograría a condición de que evitara caer en la tentación de hacer otra de sus reiteradas apologías; en este caso, la de su agente como un modelo *históricamente exitoso* y *destinado a triunfar* a nivel de los *hechos*. Sin embargo, como hemos visto, es esto último lo que precisamente termina por hacer el filósofo canadiense, a lo cual añade su afirmación de que su modelo de agente es, a todas luces, “más realista” y “aplicable” que otros (pp. 235-241, 245-257), por ejemplo, que el de Rawls. Al mismo tiempo, Gauthier reafirma su tesis de que el contexto o el ámbito de acción de su agente ideal es, fundamentalmente, el *mercado*.

Detengámonos en este último elemento de su argumentación, pues con ello nuestro autor cree que cierra exitosamente su demostración de la tesis básica de *La moral por acuerdo*. Una vez que Gauthier concluye su exposición acerca del *proviso* como el principio general que debe elegir el agente arquimédico, a continuación el

filósofo canadiense intenta establecer cuáles serían los principios derivados de este marco normativo general proporcionado por la cláusula lockeana, y termina afirmando que el resultado de aplicarla, en sus dos bases normativas de libertad y mutuo beneficio, no es otro que el *mercado*.

First, she [the ideal actor] must ensure that she, and so each person, enjoys the greatest expectation of utility compatible with the comparable expectation of every other person. Not only mutual benefit, but optimality, is the object of Archimedean choice. Given freedom constrained by the proviso, so that force and fraud are absent from interaction, a market emerges. In the circumstances of perfect competition, this market results in an optimal outcome, reached by exchanges in which each pays, and pays only the costs for the benefits he receives. [. . .] The market respects the liberty of each to choose whether or not to interact with his fellows. [. . .] The ideal for Archimedean choice is the free market society. [. . .] The market [as we saw] creates a morally free zone, but is itself a rational and moral creation. The conditions for market interaction are not always present, but their absence is never without loss for rational actors. From the Archimedean point the market is chosen as itself an ideal (pp. 261-262).

Ante esto último creo que, de nuevo, el lector bien puede preguntarse si las exigencias normativas que provienen del *proviso*, así como la imparcialidad del agente que elige la cláusula lockeana como marco general de legitimidad para las interacciones humanas, tienen que conducir, como a su resultado necesario, al *mercado* y al modelo de interacción que éste supondría; sobre todo si se quiere dar cuenta de la fundamentación de la *moral*. A estas alturas, esta tesis de Gauthier de que la sociedad capitalista y las interacciones propias del mercado constituyen el ámbito en el que se concreta el ideal normativo que se exige desde el *proviso*, pienso que bien puede suscitar, por lo menos, las siguientes dudas: ¿Qué tan ajustado estaría realmente el modelo del mercado a las exigencias de la cláusula lockeana? ¿Qué tipo de sociedad puede ser aquella cuyo modelo es el mercado y qué tan moral nos parecería? ¿Cómo se concebirían las relaciones que se establecerían en esa sociedad, si los miembros de ésta son pensados como agentes maximizadores de un mercado y, de nuevo, qué tan morales nos parecerían dichos agentes? ¿Qué aspectos importantes de las experiencias morales de las personas serían validamente modelables por el mercado y qué otros aspectos, en cambio, no tendrían cabida en ese modelo o, incluso, podrían ser opuestos a éste? Si bien Gauthier no se plantea estas inquietudes sobre las relaciones —o inconsistencias— entre mercado y moralidad, sí parece, a cambio de ello, caer en la cuenta, una vez más, de que el funcionamiento de los mercados reales haría dudar de la ejemplaridad del modelo del mercado como el *locus* en el que se realizan las exigencias normativas que

se plantean desde el *proviso*. ¿Qué sucede si, como parece ser el caso, en las sociedades reales, con mercados reales y no ideales, el funcionamiento de éstos contradice dichas exigencias de libertad y mutuo beneficio?

Tal vez previendo este tipo de inquietudes, el autor intenta poner punto final a esta discusión, ofreciendo, como su respuesta definitiva, una suerte de fusión entre los *dos polos* entre los que se ha movido su estrategia de defensa: la apelación a lo *ideal-normativo* de su modelo, y su tesis de que éste resulta ser *aplicable e histórica-empíricamente exitoso*. Así, de *una parte*, Gauthier advierte que en aquellas situaciones en las que pueda presentarse el problema de las injusticias reales, es decir, si determinado mercado real acusa fallos, lo que los actores económicos involucrados deberían hacer, es llevar a cabo los cambios que se hagan necesarios para que no haya polizones ni parásitos, ni traslado unilateral de costes, ni acaparamiento unilateral de beneficios. De este modo, dicho mercado entraría a funcionar cumpliendo con las exigencias normativas de la cláusula de Locke (p. 260). Pero me temo que esta respuesta de Gauthier, y tal vez a pesar suyo, finalmente lo que nos hace ver con mayor claridad, es que el peso *normativo* de tales exigencias está *por fuera y por encima del mercado*. Esto es, creo que la instancia normativa, en este caso representada por el *proviso*, no necesariamente ha de llevar a la conclusión de que el mercado es el modelo ideal de interacción humana justa e igualmente beneficiosa para todos. Antes bien, lo anterior tendría que llevarnos a la conclusión de que todo esquema de cooperación, incluyendo al mercado mismo –que no tendría por qué considerarse como “ejemplar” en comparación con otros esquemas posibles o reales de interacción– puede y debe ser juzgado desde una instancia normativa, v.g. el *proviso*. Esta conclusión, a mi modo de ver, tiene la importancia de que desautoriza la tesis de Gauthier de una justificación racional de la *moral* aportada desde los principios de la *teoría de la elección racional*, justificación que consiste en una suerte de “trasplante” de tales principios, desde el campo de interacción de los agentes humanos pensados como los agentes *económicos* que transan en un *mercado*, al entramado de relaciones conformado por los agentes humanos pensados como agentes *morales*.

Empero, de *otra parte*, Gauthier incluye en su estrategia una suerte de defensa del ya mencionado punto de vista arquimédico, el cual, según el autor, debería resultarnos algo bastante *familiar* (pp. 265-266), puesto que se trata de aquel punto de

vista que *cualquiera* de nosotros adoptaría cuando intente evaluar las instituciones y prácticas sociales, o cuando ensaye ponerse de acuerdo con otros agentes, a la hora de decidir la base normativa de dichas instituciones. Así, *cualquier* agente *concreto*, en el momento que lo requiera, podría hacer el ejercicio práctico de adoptar el punto de vista arquimédico y tratar de deliberar y negociar situándose en esa situación hipotética. Por ello, se ve claramente que todo sujeto *empírico* podría *identificarse* con el agente arquimédico que reflexiona de este modo.

Far from introducing a particular and artificial standpoint for choice, the Archimedean point offers the familiar perspective of agreement. A decision made in complete ignorance of one's identity proves coincident with an agreement made among persons individuated in all possible ways. [...] The Archimedean point is the position you and I occupy when we find ourselves in fully voluntary and equal agreement with our fellows. Archimedean choice yields the social contract (p. 266).

Según esto, en la familiar e intuitiva imparcialidad y universalidad del punto de vista arquimédico residiría, pues, la *aplicabilidad* del modelo que presenta Gauthier pues, tal y como él lo afirma, se trata de una instancia normativa que, lejos de ser abstracta, es aquella a la que podemos apelar todos en aquellos momentos específicos en los cuales queramos evaluar –en solitario– la justicia de las interacciones e instituciones sociales, o deliberar y discutir –con otros– acerca de la normatividad de dichas interacciones e instituciones. Sin embargo, y en franco contraste con esta posibilidad siempre abierta e intuitivamente alcanzable, nuestro autor admite que el *hecho* mismo de que determinados agentes concretos y situados en ciertos contextos reales o específicos, respeten los principios sobre los cuales dichas instituciones se legitiman, ya sería una mera cuestión *fáctica* que no puede estar asegurada de antemano por el carácter imparcial, ni por la cualificación moral del proceso de decisión, el cual, se supone, está llamado a producir o legitimar esos principios. En otros términos, el punto de vista arquimédico puede que legitime –como instancia normativa– dicho proceso y los principios acordados mediante él, pero por más que podamos acceder a él de manera intuitiva, no puede garantizar el evento concreto de que determinados agentes racionales situados en ciertas circunstancias específicas, como aquellas en las cuales nos movemos todos en el día a día, vayamos de hecho a respetar el *proviso* y el contrato social. Así, la voluntad de cumplir con el acuerdo y la motivación para ello no estarían aseguradas por la elección arquimédica de los principios de justicia.

However, Archimedean choice does not itself yield compliance with the social contract. Although each person identifies with the ideal actor's choice of social principles, he is not, thereby rationally committed to personal compliance with those principles (p. 266).

En este punto, Gauthier nos recuerda lo que arriba ya se comentó con respecto de la elección *condicional* de la propia inclinación a cumplir o no cumplir con el pacto social, dependiendo de las circunstancias en las que uno se encuentre: no sería racional comportarse como un MR en medio de un grupo de MMII, ni tampoco ser MI en medio de un grupo de MMRR. El agente ideal, dado que es pensado en condiciones ideales, no elige sin más la disposición a cumplir y ser un MR, ni tampoco la de no cumplir y ser un MI, sino que más bien elige los principios que deben guiar las instituciones y las interacciones de la sociedad que intenta diseñar. Aquí de nuevo el autor intenta responder a las críticas provenientes de sectores a los que califica de marxistas y feministas, los cuales le señalarían la poca pertinencia y aplicabilidad que su modelo tendría para contextos reales. Según dichas críticas, desde ese modelo se asumen unas condiciones de igual libertad y de mutuo beneficio como algo que puede ser garantizado por un marco legal o *formal*, cuando el caso es que dichas condiciones formales pueden ser y, de hecho, han sido contradichas históricamente por los resultados *concretos* y vividos del proceso de socialización de los agentes (pp. 321-323, 352-353). Por ejemplo, en dicho proceso pueden haber jugado un papel fundamental desigualdades tales como aquellas asociadas al género, al color de la piel y al origen social. Así aunque v.g. en muchos escenarios sociales reales hombres y mujeres *formalmente* sean considerados como “iguales” ante la ley, de todos modos, ello no obsta para que a los primeros *de hecho* se les proporcionen –por más que la ley rece lo contrario– los medios concretos para desarrollar talentos y tener oportunidades que, por el contrario, se hacen difícilmente accesibles a las segundas.

En conclusión: si se asume esta crítica, se verá que los principios de justicia elegidos en el punto de vista arquimédico bien pueden terminar por no incidir en los aspectos más sustantivos y determinantes de la convivencia efectiva entre las personas y comunidades reales. Ante esta objeción, Gauthier responde, y creo yo que en franco contraste con su anterior llamado al carácter *ideal* de su agente arquimédico, que dicho agente no sólo elige los principios de justicia enunciados en la cláusula lockeana, sino

que también elige el *proceso de socialización* y la afectividad o los *sentimientos morales* de los agentes que se guiarán por esos principios de justicia.

The ideal actor chooses not compliance, but those processes of socialization, that promote the circumstances in which narrow compliance is rational. [. . .] Each person has reason to prefer that everyone has affectively engaged by compliance, so that the familiar feelings of respect and resentment, of self-respect and guilt, are linked appropriately with the fair and unfair behaviour of others and oneself. Although primary concern is with the principles that would be chosen from the Archimedean point, we should not forget the importance of the choice of affections, in so far as these can be shaped by socialization (p. 266).

En esta parte de su texto creo que Gauthier incluye un elemento más *sustantivo* pero, por ello mismo, pienso que también bastante difícil de encajar en los esquemas que hasta ahora el autor ha presentado, los cuales se referían sólo a los aspectos puramente *formales* del diseño institucional de la sociedad. En contraste con ello, este nuevo elemento —más cercano a filosofías de cuño humeano, pero también, si se quiere, aristotélico y, por supuesto, al feminismo y al marxismo frente a los que Gauthier se ha mostrado tan crítico—, apunta hacia terrenos que nuestro autor ya anteriormente ha invalidado como tierras vedadas. Y vedadas no sólo para el tipo de investigación que él se propone sino, en general, también para *todo* discurso filosófico que pretenda ocuparse en serio de una fundamentación propiamente *racional* de la moral, o de la moral como un asunto racionalmente abordable. Con este nuevo movimiento en su línea de argumentación, repito, Gauthier podría estar traicionando su anterior estrategia de apelar al carácter *ideal* y normativo de su modelo, carácter que, como él mismo parecía suponer anteriormente, le inmunizaría frente a aquel tipo de objeciones que intentarían desautorizarlo como ingenuo, poco realista, mistificador o ideologizante. Creo que ahora la apelación a asuntos tales como los *sentimientos morales* y los *procesos de socialización* aparece como un salto no justificado e, incluso, contradictorio con la anterior estrategia argumentativa desplegada por nuestro autor. Para que este salto no se viera como tal, pienso que merecería por lo menos una explicación que permitiera desvanecer la sospecha de una posible inconsistencia. Sin embargo, tal explicación no aparece y, aún así, con este último paso Gauthier cree que logra cerrar concluyentemente su demostración de que moralidad y elección maximizadora no sólo no se contradicen entre sí, sino que, además, se solicitan mutuamente.

Moral theory offers an Archimedean point analysis of human interaction. The theory of rational choice offers an analysis from the standpoint of each interacting individual. In exhibiting the harmony between these analyses we have shown that the moral demand for impartiality can be accommodated within the rational demand for individual utility-maximization, by attending to the conditions of agreed interaction among equally rational persons (pp. 266-267).

No obstante lo anterior, el mencionado “salto” que presenta la argumentación de Gauthier hacia terrenos que, según las ideas iniciales de su texto, serían ajenos tanto a su modelo de racionalidad, como a su noción de moral, esto es, terrenos en los que hallamos tematizados asuntos tales como los procesos de socialización y los sentimientos morales, creo que en *La moral por acuerdo* admitiría ser visto como un paso que cumple un importante papel. A saber, el de ir allanando el camino para un giro fundamental dentro del desarrollo del texto: la conversión del *homo oeconomicus* en el individuo liberal, figura con la cual nuestro autor anuncia que se comprometerá definitivamente como el agente modélico de su discurso.

Este individuo liberal, como veremos a continuación en el siguiente capítulo, al ser un personaje en el que se incluyen elementos más cercanos a las intuiciones morales de discursos antes rechazados por Gauthier, elementos que superan los cálculos del maximizador, a los ojos del autor permitirá que se salven las dificultades impuestas por la figura del *homo oeconomicus*, dificultades que terminarían, como hemos visto, por hacer dudar de la auténtica motivación moral del agente propuesto por Gauthier. Dicha duda, como intentará mostrarse, traería la importante consecuencia de que su fundamentación de la moral terminaría por ser vulnerable a la sospecha escéptica. Veamos, entonces, a continuación, cómo tiene lugar la transformación del *homo oeconomicus* en el individuo liberal y, en estrecha relación con esto, cómo responde el autor frente a aquellas críticas que podrían acusarle de presentar un agente moral cuyo compromiso con las normas fuese tan débil –al estar atado únicamente a su autointerés– que pondría en peligro la idea misma de una moral. Con lo cual, no habría más remedio que ver en Gauthier un escéptico y en su moral por acuerdo un discurso incapaz de competir con otras propuestas filosóficas que intentan dar cuenta de la moralidad desde posiciones cognitivistas; resultado al que nuestro autor intenta escapar, mediante la mencionada transformación en su modelo de agente.

Capítulo 3

La conclusión del proceso de conversión del *homo oeconomicus* en el individuo liberal

Ante las mencionadas sospechas que, como se vio en el capítulo anterior, podrían recaer sobre la *moralidad* atribuible tanto a una forma de cooperación basada en el *cálculo maximizador*, como a los agentes *egoístas* que tomen parte en ella, en los últimos dos capítulos de *La moral por acuerdo* Gauthier intenta demostrar que su propuesta logra encarar con éxito dicha sospechas. Nuestro autor se hace cargo de que si, como él lo sostiene, la moral racionalmente justificada sólo debe pensarse como el producto de un pacto entre maximizadores egoístas, entonces no faltarían las voces críticas que podrían descalificar como amoral la motivación atribuible a este tipo de agentes; la posible inestabilidad y la falta de fuerza vinculante que suele caracterizar los acuerdos suscritos en razón de la conveniencia y/o del egoísmo, así como el peligro de que, por estas mismas causas, una moral de este tipo bien pueda ser utilizada como una mera máscara. Esto es, como un sofisticado instrumento de dominación del que los astutos maximizadores echarían mano cuando necesiten obtener de un modo más eficiente (*i.e.* menos costoso, al ser menos violento), aquello que también podrían lograr mediante el uso de la fuerza.

Si este fuese el caso y la moral no fuese más que una fachada para ocultar y favorecer los intereses de ciertos agentes, posiblemente en contra de los intereses de otros, no sería sostenible la tesis de Gauthier de que hay un auténtico vínculo entre la moral de los maximizadores y la racionalidad humana, a menos que en *La moral por acuerdo* se haya cometido, desde su planteamiento inicial, una *petitio principii*. Nuestro autor se enfrenta, entonces, a la necesidad de mostrar que su teoría no conduce en absoluto a este devastador resultado, aunque, al mismo tiempo, insista en defender el modelo maximizador como la legítima base racional para la moralidad y, por tanto, también el modelo de los agentes egoístas como el paradigma de la agencia humana moral-racional. A continuación examinaremos una ingeniosa estrategia utilizada por Gauthier para desautorizar estas sospechas que tanto teme y que, como digo, amenazan

con dar al traste con su propuesta: el filósofo canadiense intenta refutar algunos de los viejos –pero aún preocupantemente actuales– argumentos de los sofistas, tal y como aparecen en los textos platónicos.

1. Las sospechas del sofista: la justicia concebida como “mal necesario”

En el capítulo X de *Morals by Agreement* (pp. 306 y ss.) Gauthier se vale del mito del anillo de Giges, que Platón pone en boca de Glaucón en la *República* (358e-360e), con el objetivo de ilustrar la tesis, defendida por Trasímaco (en el mismo texto), de que la justicia es sólo un mal necesario al que acuden los débiles, dado que no son lo suficientemente poderosos para dominar a sus congéneres e infringirles injusticia sin sufrir retaliaciones. La situación de aquellos que no pueden dominar del todo a los demás es también la de quienes de alguna forma dependen aún de ellos para satisfacer sus necesidades o para no sufrir perjuicios. De allí que hacer un pacto de no agresión mutua sea para algunos, según el relato del sofista⁹⁷, un recurso nacido de la debilidad humana, de la menesterosidad que nos caracteriza a los unos frente a los otros y que a todos nos pone en riesgo, así como de la necesidad que tenemos de sobrevivir en dichas condiciones de menesterosidad y peligro. Siempre existirá la posibilidad de que nos infrinjamos daño, pero no tenemos más remedio que acudir a los otros, ante el hecho innegable de que precisamos su ayuda. Así, según cuenta Glaucón haciendo eco de la tesis de Trasímaco, la justicia es vista por muchos como un término medio entre el mayor bien, que consistiría en poder ejercer la injusticia contra otros sin recibir castigos a cambio, y el mayor mal, que estaría en padecer la injusticia por parte de los demás sin recibir una compensación o sin tener la posibilidad de la revancha.

A partir de la narración del mito del anillo, el sofista intenta obtener una conclusión escéptica en contra del fundamento racional que Sócrates y otros parecen atribuirle a la justicia entre los hombres, conclusión que tiene la importancia –entonces para Platón y ahora para Gauthier– de que, como ya se ha anunciado, con ella se niega el vínculo que tanto le interesa demostrar al filósofo canadiense: el nexo entre razón y moralidad. Mediante el mito, Glaucón ilustra la hipotética situación de alguien que

⁹⁷ Si bien Glaucón advierte (retóricamente) que él sólo se limita a contar lo que otros dicen (358c-358d).

tuviese el poder para infringir injusticias a los demás sin que éstos pudiesen penalizarlo por ello, pues el anillo, al volverle invisible, le escondería de los otros, haciéndole inaccesible para ellos y situándole, por tanto, por encima de los demás hombres y a salvo de la necesidad de pactar con éstos. Gracias a este maravilloso poder que le otorga el anillo, el trasgresor obtendría todo lo que quisiera, sin necesidad de incluirse en la sociedad ni, por tanto, de respetar sus leyes. Esta es exactamente la realidad opuesta a la de aquellos que, como la mayoría de nosotros, no tenemos más remedio que ceder ante nuestros congéneres y llegar a acuerdos con ellos, puesto que aún somos vulnerables a sus reacciones y además, podríamos necesitar de su ayuda para suplir nuestras necesidades. Pero aquel que tuviese un poder como el que concede el anillo, según el sofista, estaría por fuera de sus cabales –i.e. sería irracional– si firmara pactos de no agresión con sus semejantes, pactos en los que se comprometiera a cumplir un código común de normas de justicia, ya que no tendría ninguna razón ni para temer a los demás ni para solicitar su colaboración.

La justicia, pues –y esta es la conclusión que tanto preocupaba a Sócrates y Platón, y ahora preocupa a Gauthier–, no tendría por qué ser pensada como una necesidad *racionalmente* dictaminada. Por el contrario, más bien tendríamos que considerarla un mal necesario; un recurso al que acudimos convencionalmente, en vista de nuestra débil condición humana. De modo que si se diese el caso de que, por alguna circunstancia, se nos liberara del yugo que dicha debilidad nos impone, entonces también nos veríamos libres de las razones que nos llevaron a respetar la justicia y no abusar de las otras personas. ¿Qué razones son esas? Volvamos a la pintura de ese ser que se hace todopoderoso gracias a los efectos del anillo, pintura que, repito, es la contraria a la de la condición humana “normal” pues, amén de que no suele ocurrirnos que logremos dominar del todo a nuestros semejantes, preciamos de su buena voluntad con el fin de asegurarnos su ayuda para suplir nuestras necesidades y carencias. De allí que nuestra experiencia sea la de la mutua dependencia y la humana vulnerabilidad, experiencia que nos ha llevado a hacernos conscientes de lo desagradable que es sufrir por falta de auxilio, o a causa de los males o las injusticias que otros nos infrinjan. Esta experiencia nos hace caer en cuenta de que los placeres o los beneficios que se puedan obtener haciéndoles daño o siendo injustos con ellos, son excedidos en mucho por los sufrimientos y perjuicios que se padecen cuando uno(a) mismo(a) es víctima de la

insolidaridad o, peor, del mal y de la injusticia que los demás le infrinjan. Igualmente, nos percatamos de lo terrible que puede ser recibir los castigos o sufrir las retaliaciones por parte de quienes han sido objeto de nuestras agresiones o injusticias, o de nuestra negativa a ayudarles.

De manera que, por esta vía, llegamos al convencimiento de que lo más razonable que podemos hacer los seres humanos, ante nuestra ineludible fragilidad y de que no está a nuestro alcance el ejercer un poder sin límites sobre nuestros congéneres, es pactar y comprometernos a no ser injustos los unos con los otros; o por lo menos, fingir que hemos sido sinceros al suscribir tal pacto y, así, no despertar la desconfianza de las demás personas. Esta es, pues, la salida racional de quienes están en la condición humana “normal” pero, obviamente, según el sofista, estaría claro que no se trata de la opción más razonable para un individuo que sí pudiera ejercer un poder absoluto sobre sus semejantes: en su caso sería una insensatez pactar y comprometerse a ser justo. Esta misma tesis que, insisto, niega el vínculo entre razón y moralidad, de suyo le plantearía un serio problema a todo aquel que intente fundamentar racionalmente las normas morales y justificar como auténticamente *moral* el tipo de *motivación* que lleva a un agente *racional* a comprometerse con dichas normas. De allí que el filósofo canadiense reactualice la vieja cuestión sofística preguntándose (p. 307): si un agente pudiera asegurarse los bienes que necesite sin contar para ello con la ayuda de nadie, y si tuviese el poder de perjudicar a otros sin la perspectiva de recibir a cambio un castigo, ¿qué razones podrían convencerle de obedecer a la moral? ¿Acaso no sería un insensato si, estando en tan privilegiada situación, respetara las restricciones que la moral le impone a la búsqueda de su propio beneficio? Si así fuera, ¿no sería, entonces, absurdo o, en el mejor de los casos, un acto de hipocresía insistir en hablar de una moral “racional”?⁹⁸

De hecho, en el texto platónico, el sofista llega a afirmar que no es poco frecuente que los poderosos, cuando se les presenta la ocasión, utilicen astutamente este recurso de la hipocresía, al enmascarar sus auténticas intenciones. De este modo convencen a los más débiles de que los primeros también creen sinceramente en la justicia de las normas sociales y, por tanto, así mismo se comprometen a respetarlas en serio, por más que el poder que detentan gracias a su mayor estatus en la jerarquía social eventualmente les permitiría situarse por encima de la ley. Con ello se aseguran la

⁹⁸ Salvo mi paráfrasis y mi traducción.

docilidad de los menos favorecidos, aprovechándose de su inocencia y debilidad, o sacando partido de aquellas circunstancias que impiden a la mayoría de los hombres ejercer un poder sin límites. El lector habrá advertido que si Gauthier busca ofrecer una fundamentación racional de la moral, debe entonces, mostrar las razones que habría para no estar de acuerdo con esta tesis que Platón pone en boca de Trasímaco y Glaucón y, de paso, evitar que su propuesta sea asimilada a dicha tesis. Es decir, tendría que mostrar las razones para considerar a la moral como algo positivo que está lejos de ser un mal necesario al que nos resignemos dada nuestra débil condición humana o, mucho menos, un conveniente engaño del que se valen los poderosos para conservar, en favor suyo, la paz social.

1.1 La idea de un hombre “natural” y pre-moral

Por lo anterior, el autor canadiense se encuentra en una incómoda posición frente al hecho de que, a partir de los relatos del sofista y también de algunos autores modernos, como Hobbes,⁹⁹ la moral podría ser vista como una imposición sobre nuestra “verdadera naturaleza”; como una necesidad con la que nos encontramos luego de que, repito, hemos sufrido injusticias por parte de nuestros semejantes y por eso llegamos a ver la conveniencia de ese mal menor que son las restricciones morales. Insisto en que la posición de Gauthier no es nada cómoda pues, de *un lado*, y como ya se comentó en el primer capítulo, se le puede acusar de suscribir la idea de que la moral bien puede ser considerada como una suerte de “invento”, de artefacto social que se superpone a nuestra “auténtica” naturaleza siendo, pues, un artificio que nos constriñe en nuestra natural búsqueda del beneficio personal. Empero de *otra parte*, nuestro autor *no* desea en absoluto –y es aquí donde su posición se torna un tanto difícil de sostener– darle la razón a Trasímaco cuando, amén de lo anterior, este último afirma que la moral podría ser utilizada por los poderosos como una conveniente máscara que les permita asegurar la lealtad o docilidad de los más débiles e inocentes.

Según Gauthier, si nos tomamos en serio la tesis sofista (o hobbesiana) de una humanidad “original” o “natural” que es constreñida por ese artefacto social que serían las normas morales, entonces el ser humano “original” tendría que ser considerado

⁹⁹ Por cierto, Gauthier no menciona a Nietzsche.

como alguien que de suyo es amoral; que nace y bien puede situarse —si las circunstancias se lo permiten— por fuera de las restricciones morales. Vistas así las cosas, en la “naturaleza” que, entonces, tendría que ser pensada como previa al pacto social, no podría hallarse injusticia ni tampoco justicia alguna. El hombre “pre-social” o “natural” sería para Hobbes y para los sofistas, según lo afirma nuestro autor, un ser pre-moral que acude a la moralidad únicamente llevado por la fuerza de sus necesidades. ¿Está o no de acuerdo Gauthier con esto último? Creo que ya el lector, como le ocurre al autor de *La moral por acuerdo*, se habrá hecho cargo del enorme parecido entre su propia tesis y aquella que intenta ahora desautorizar, y que pone en boca de los sofistas y de Hobbes. De modo que si Gauthier se propone objetar a estos últimos y defender su propia propuesta debe, entonces, acometer la difícil tarea de mostrar que *su* idea de la moral *no* es equiparable en absoluto ni a la patraña que cree desenmascarar Trasímaco, ni tampoco al mencionado mal necesario al que debemos resignarnos cuando no somos lo suficientemente poderosos para dominar a nuestros semejantes.

Para lograr este doble y acaso contradictorio cometido, la moral que postula el filósofo canadiense debe mostrarse como inmune a toda crítica que intente, o bien suscribir la postura sofista, o bien desautorizarla apelando a razones a las que Gauthier antes ha descalificado como “neo-kantianas”. Sólo si logra ambos propósitos nuestro autor podría sostener con mayor éxito la plausibilidad de su idea de una moral racional que, empero, no sea “racional” en el mismo sentido en que lo sería para Trasímaco. De esta manera, también sería capaz de superar el escollo representado por los argumentos del sofista en su discusión contra Sócrates. No obstante, para lograr este doble fin Gauthier debe sortear, por lo pronto, tres obstáculos. Como ya se ha mencionado, en *primer* lugar nuestro autor está firmemente comprometido con la idea de la racionalidad humana como exclusivamente *maximizadora*. En *segundo* término el filósofo canadiense ha insistido en atribuir una psicología egoísta a sus maximizadores *no-concernidos* por la suerte de sus semejantes. Y *finalmente* habría un inconveniente que, como se verá, reviste una importancia especial. Me refiero al hecho de que la propia tesis de Gauthier se apoya en una idea a la que, paradójicamente, también acuden el tipo de discursos de los que intenta desmarcarse nuestro autor: la noción de un “hombre natural” o “pre-social”.

Gauthier se percata entonces de que, si se hace uso de esta figura, no sólo puede llegarse a los resultados por él buscados sino que, en contra de lo que a él le parece correcto, también podrían mostrarse serias razones para dudar de una auténtica motivación moral en el caso de los maximizadores de *La moral por acuerdo*. De modo que el canadiense se encuentra con la contradictoria necesidad de tener que defender la idea de un hombre “natural” y “pre-social”, al tiempo que debe proteger su propio discurso frente a los ataques que, como acabamos de ver, podrían desautorizarlo apelando –justamente– a esa misma idea. Con el fin de efectuar este difícil movimiento en su argumentación, Gauthier intenta responder a una conocida crítica en contra de esta figura un ser humano pre-social: la de A. MacIntyre. El autor canadiense (pp. 310 y ss.) ve que los dardos lanzados por el autor de *Tras la virtud* también pueden hacer mella en su tesis de que puede haber una auténtica moral entre maximizadores autointeresados y no-concernidos por la suerte de sus congéneres; sobre todo si se tiene en cuenta la vulnerabilidad que, frente a objeciones como las de MacIntyre, presentaría el modelo del agente racional-moral tal y como aparece presentado en la obra de Gauthier. Es decir, como un *homo oeconomicus* y, en estrecha conexión con esta noción, la idea de que la comunidad moral ha de ser pensada como una construcción posterior al contrato social, o como una red de agentes económicos que se instrumentalizan los unos a los otros: el mercado. En su *Historia de la Ética*, MacIntyre descalifica tanto la noción sofista de un “hombre natural” (pp. 26-28), como la versión moderna de esta idea (pp. 44, 54-55, 132, 158), claramente expuesta en autores como Hobbes.

La razón de esta descalificación está en que, para MacIntyre, se trata de una noción incoherente, puesto que con ella se supone que ha existido algo así como un ser humano anterior a toda historia social y, sin embargo, subrepticamente esta misma idea se corresponde con una visión del hombre que ha sido históricamente construida y que, por tanto, no representa en absoluto, aunque así se lo intente mostrar, a un ser humano que está por fuera o por encima de los avatares de la historia, o de las características de una sociedad humana específica y empíricamente investigable. En el caso de la noción expuesta por Trasímaco y Glaucón, se da por sentado que ha habido un hombre “natural” previo a todo pacto social por el cual los hombres han decidido respetar la moral y las leyes pero, en realidad, dice MacIntyre, este concepto se construye con base en la imagen homérica del héroe que antecede el nacimiento de la *polis*. Esto es, un

hombre que pertenece a una época concreta que tiene sus propias normas, no un ser que esté por fuera de los tiempos ni de las reglas sociales. Según el autor escocés, esta misma trampa se esconde, con las adaptaciones del caso, tanto en la noción moderna del hombre “natural” y racional que es previo al contrato social o que antecede a la moral y a la justicia, como en la idea conexas a ésta: la de un supuesto estado de naturaleza en el que cada cual se procura –racional, *i.e.* egoístamente– su propio beneficio (y si, es el caso, sólo el de sus seres queridos), aún a costa de dañar a los demás hombres.

MacIntyre claims that Western moral philosophy has repeatedly committed the sophistic error of supposing that there is an extra-moral standpoint, natural or rational from which morality can be grasped and either defended or attacked (p. 310).

Ante las consecuencias indeseables que esta crítica de MacIntyre, como bien puede advertir el lector, también traería en contra de la propia tesis de Gauthier, éste ensaya una interesante y astuta estrategia argumentativa: intentará mostrar que esta noción de un hombre “natural”, racional y pre-moral *no* es de suyo incoherente ni tramposa, sino que se trata más bien de una ficción necesaria y muy conveniente para abordar el tema de la *motivación* moral. Nuestro autor expresamente afirma que él reconoce, con MacIntyre, el hecho innegable de que el hombre nace y vive en una comunidad lingüística viéndose, desde el inicio de su proceso de socialización, involucrado en las convenciones de su cultura, las cuales le moldean como persona. Sin embargo, Gauthier cree que ello no obsta para que aún quepa preguntarse si esas convenciones introducen de manera imprescindible –definitiva e insoslayable– en las motivaciones y valores de todo individuo, el elemento de sociabilidad; es decir, aquello a lo nuestro autor se refiere como “tuismo” o valores y actitudes “tuistas”.¹⁰⁰ Ante esta cuestión Gauthier cree que la respuesta es negativa: nuestro no-tuismo no necesariamente resulta ser neutralizado de manera radical o definitiva por nuestro tuista proceso de socialización. Por ello sigue siendo preciso investigar las razones por las cuales renunciamos a ese no-tuismo y decidimos seguir reglas morales de carácter tuista, o nos motivamos a actuar moralmente.

¹⁰⁰ Remito al lector a la nota de pie de página (en el primer capítulo) en la que antes advertí que no tengo claro cómo traducir *tuistic* ni *tuism* y, por ello, me tomo la licencia de castellanizar de mala manera estos términos, que en inglés posiblemente no ofrezcan mayores problemas, pero que en español no estarían reconocidos en los diccionarios oficiales.

Our question is then whether the conventions that enter into the make-up of a person must introduce any element of sociability into either her valuation or her motivations, and it seems clear that they need not. [. . .] The condition for an asocial valuation is simply non-tuism; the non-tuistic takes no interest in the interests of those with whom he interacts . . . But non-tuism does not ensure social motivation. Even if one takes no interest in other's interests, one may take an interest in states of affairs than cannot be specified except with reference to others. [. . .] Thus in accounting for the motivation of person X, we may refer only to those features of states of affairs that involve X and no one else (pp. 310-311).

1.2 Para salvar al hombre natural: motivaciones “primarias” vs. “secundarias”. De nuevo, el *homo oeconomicus* como talón de Aquiles de la propuesta de Gauthier

Nuestro autor cree, entonces, hallar la manera de rescatar la idea sofista y hobbesiana de un sujeto “puro” o “natural” que puede salirse de su propia historia y decidir si en realidad quiere o no hacer caso de la moral y, por tanto, si aceptaría prescindir o no de los lazos –morales– que le atan con otros agentes, aunque con esto, repito, Gauthier no quiere llegar a la conclusión negativa de que la moral es un mal necesario o un instrumento de dominación. Para ello el filósofo canadiense intenta mostrar por qué no debe hacerse caso de una descalificación como la que MacIntyre hace del punto de vista extra-moral y pre-moral, al considerarlo sólo un artificio, una idea tramposa e incoherente. Para Gauthier esta descalificación nace de tomarse demasiado en serio la idea tuista del ser humano como ser socializado, idea desde la cual, a pesar de lo que se afirme desde posiciones como las de MacIntyre, no puede mostrarse, según el filósofo canadiense, que su contraria, es decir, la noción no-tuista del hombre pre-moral, sea incoherente. Por el contrario: si bien, y como ya se ha mencionado, Gauthier acepta el innegable hecho de que somos seres convencionalizados, también insiste en que la idea no-tuista resulta sumamente útil, amén de plausible, para poder reflexionar acerca de qué estaría tras la –acaso sólo aparente– motivación moral que parecen tener muchos agentes para el tipo de acciones que ellos suelen calificar de “justas” o “decentes”.

Además de su utilidad conceptual, para nuestro autor la idea no-tuista del hombre “natural” tiene también un innegable anclaje en la realidad, sobre todo, si se toman en cuenta las anteriormente comentadas circunstancias de la escasez y del

autointerés que caracterizan la condición humana. Recordemos que, según Gauthier, en un mundo de recursos escasos no hay más remedio que competir y buscar el propio beneficio, y en ello no le faltaba razón a Hobbes. De allí que, una vez más, obedeciendo a su técnica de introducir díadas dicotómicas, el filósofo canadiense defina dos tipos básicos de motivaciones humanas. En *primer* lugar existen, según él, aquellas a las que se refiere como “motivaciones primarias” y que para nuestro autor son las que nacen directamente de las circunstancias de la escasez y la necesidad de competir. En *segundo* término estarían las motivaciones a las que Gauthier llama “secundarias”, las cuales, según él, se superponen a las primeras y vienen determinadas por otra necesidad: la de establecer lazos sociales (pp. 310-311).

De manera que, para el autor de *La moral por acuerdo*, habría una asociabilidad subyacente a los comportamientos sociales, lo cual en su concepto hace ver que la lectura correcta que tendría que hacerse de la idea de un hombre “natural” no debe ser en absoluto aquella que hace MacIntyre. Esto es, la noción que tanto cuestiona el filósofo escocés no debe ser considerada incoherente ni tramposa, sino que más bien habría que verla como un recurso conceptual que evidencia una innegable tensión entre dos aspectos distintos pero coexistentes en una única naturaleza humana, aspectos que orientan nuestros actos y motivaciones en direcciones opuestas, pero a las que tiende indudablemente cada individuo, y sin que ello signifique contradicción alguna.¹⁰¹ En todo caso, para Gauthier, la tan cuestionada idea nos permite llegar a la conclusión, bastante plausible, de que la naturaleza de la sociabilidad humana es convencional y,

¹⁰¹ Compárese con las ideas kantianas de la “insociable sociabilidad” y de la “madera torcida” de la humanidad (Trad. 2006, pp. 8-12): “El medio con que se sirve la Naturaleza para llevar a cabo el desarrollo de todas sus disposiciones es el antagonismo de las mismas dentro de la sociedad, en la medida en que ese antagonismo acaba por convertirse en la causa de un orden legal de aquellas disposiciones. Entiendo aquí por antagonismo la insociable sociabilidad de los hombres, esto es, el que su inclinación a vivir en sociedad sea inseparable de una hostilidad que amenaza constantemente con disolver esa sociedad [. . .] Esta necesidad que constriñe al hombre –tan apasionado por la libertad sin ataduras– a ingresar a ese estado de coerción, es en verdad la mayor de todas, esto es, aquella que se inflingen mutuamente los hombres, cuyas inclinaciones hacen que no puedan coexistir durante mucho tiempo en salvaje libertad. Sólo en el terreno acotado de la asociación civil esas mismas inclinaciones producirán el mejor resultado: tal y como los árboles logran en medio del bosque un bello y recto crecimiento, precisamente porque cada uno intenta privarle al otro el aire y el sol, obligándose mutuamente a buscar ambas cosas por encima de sí, en lugar de crecer atrofiados y encorvados como aquellos que extienden caprichosamente sus ramas en libertad y apartados de los otros. [. . .] El jefe supremo debe ser, sin embargo, justo *por sí mismo* sin dejar de ser un *hombre*. Por eso esta tarea es la más difícil de todas y su solución perfecta es poco menos que imposible: a partir de una madera tan retorcida no puede tallarse nada enteramente recto. La naturaleza sólo nos ha impuesto la aproximación a esa idea” (Ak., VIII, pp. 19-23).

por ello mismo, en el intento de dar cuenta de dicha sociabilidad y de qué es aquello que nos *motiva* a buscarla o conservarla, no tiene por qué extrañarnos e, incluso, resulta inevitable, el que arribemos a la noción de un hombre “natural” o pre-moral.

. . . So [that] what is revealed when all social practices and institutions are removed is a being asocial in motivation and non-tuistic in values (p. 311).

De manera que, para nuestro autor, es importante rescatar esta idea de un agente que puede situarse por fuera de la sociedad y cuestionarla, no como si fuese un *hecho* el que un individuo pueda vivir o haya vivido por fuera de esta última, sino como un recurso conceptual necesario para que pueda establecerse la naturaleza de la *motivación* que nos mueve a participar en dicha sociedad y cumplir con los mandatos morales. En este punto creo que el lector bien puede señalarle a Gauthier, como seguramente también lo haría MacIntyre, que ese retrato de un sujeto pre-moral y extra-moral, no sólo admite ser visto como un producto conceptual desinteresado, propio de quien asume (dicho en una suerte de jerga hegeliana) el “esfuerzo del concepto” y trata de enfrentarse cándidamente a los problemas filosóficos, reflexionando sin ataduras históricas previas. Pienso que esta idea también podría ser considerada como una construcción que le debemos al tipo de sociedad en la que hemos aprendido a concebirnos a nosotros mismos y a convivir con otros seres humanos. Pues en esta sociedad que, repito, es la que nos ha sobrevenido históricamente, tal vez se nos haga cada vez más clara la posibilidad de pensarnos y de comportarnos como seres desvinculados, para los cuales sus lazos sociales son sólo un medio, y un medio para fines no-sociales. En razón de ello, puede que muchos de nosotros hayamos terminado por considerar a la sociedad como un aparato que conformamos de manera posterior –y voluntaria y estratégicamente– a nuestro mutuo encuentro como individuos; por ende, acabemos por pensarnos a nosotros mismo como si hubiésemos sido constituidos, en tanto que tales individuos, aún antes de tal conformación de la trama social. De allí que esa sociedad sea algo que nos hemos acostumbrado a asumir como un artefacto del que podríamos haber hecho caso omiso, o del que seríamos capaces de prescindir siempre que lo encontremos necesario.

Por lo anterior, creo que tal vez algo o mucho de razón lleve MacIntyre cuando acusa a una importante corriente de la filosofía moderna de haberse tomado demasiado

en serio lo que podría ser sólo una ficción útil para pensar problemas; y tan en serio que ha podido llevarnos al equívoco de asumir que es así como *somos “esencial”, “natural” o “necesariamente”* los agentes humanos; esto es, seres pre-sociales que se deciden a cooperar entre sí únicamente llevados por intereses no-sociales. Amén de que, repito, autores como Gauthier no parecen reparar demasiado en la sospecha de que esta visión de nosotros mismos puede hacer parte de una forma *históricamente rastreada* y, por tanto, también *epistemológica y moralmente falible*, tanto de concebirnos a nosotros y a nuestros congéneres, como de establecer, explicar y legitimar nuestras relaciones con ellos. Como hemos visto, de su pintura no-tuista de los agentes morales Gauthier busca rescatar su conexión con lo fundamental de su propuesta de una moral por acuerdo, con el fin de que ésta pueda resistir los ataques provenientes, de un lado, de posiciones como la de MacIntyre y, de otra parte, de posturas escépticas, tales como la que ejemplifican los sofistas. Como parte de su defensa frente a este tipo de escepticismo, arriba se comentó que nuestro autor intenta mostrar que la moralidad no debe ser vista como un mal necesario nacido de nuestra debilidad y, lo que es más importante, tampoco debe arribarse a la devastadora conclusión de que las normas morales son sólo un instrumento de dominación –ideológica– de los poderosos sobre los débiles, o de aquellos ricos que hacen lo posible por mantenerse a salvo de los pobres.

Esta última afirmación también es atribuida por nuestro autor (pp. 312 y ss.) a Rousseau, sobre todo en su *Discurso sobre los orígenes y los fundamentos de la desigualdad entre los hombres*.¹⁰² Si esta acusación del autor ginebrino afecta igualmente a la filosofía moral contractualista en general y, por ende, también a la moral por acuerdo propuesta por Gauthier, éste intentará mostrar que los pobres y los débiles también tendrían razones de peso, al igual que los ricos y poderosos, para acatar los mandatos morales. Para ello recuerda, en primer lugar, su ya comentada tesis de la “igualdad en la racionalidad” y, en conexión con ella, su idea de que la justicia y la razón coinciden. Así, Gauthier repite lo anteriormente dicho: si llegara a ocurrir que en algunos escenarios reales no parezca ser el caso que todos los agentes involucrados puedan ser igualmente racionales y, por tanto, que racionalidad y moral no coincidan, o

¹⁰² Cf. Rousseau (Trad. 2001, *Segundo Discurso*, II, pp. 178-183). En este punto Gauthier no menciona a Nietzsche, Marx, Foucault, ni a nadie cercano a la así llamada “filosofía de la sospecha”, ni tampoco al feminismo.

que en dichas circunstancias lo razonable fuese comportarse como un maximizador irrestricto MI, nuestro autor también cree que se trata de meras circunstancias accidentales y superables. De este modo, tales circunstancias pueden ser corregidas “igualando” la racionalidad de las partes implicadas (pp. 313-314)¹⁰³ y aplicando además, como remedio, el *proviso* lockeano. Una vez implementados estos correctivos y superadas las diferencias en cuanto al nivel de racionalidad de las partes, entonces sí debe ocurrir, según el filósofo canadiense, que necesariamente racionalidad y justicia coincidan. De este modo ya no vendrían al caso los señalamientos de los sofistas y de autores como Rousseau cuando sostienen que la sociedad permanece estable sencillamente porque los débiles o los pobres –en su inocencia y vulnerabilidad– cumplen con su parte del contrato social, mientras que los poderosos o los ricos sólo simulan hacerlo.

Creo que a Gauthier lo que le preocupa es la perturbadora tesis escéptica de que estos últimos, si en algunas ocasiones llegaran a necesitar hacer caso omiso de las restricciones morales, no dudarían en hacerlo dado que, por una parte, y en razón de su poder, de su carencia de limitaciones, están en capacidad de transgredir los mandatos morales sin recibir castigos a cambio. Por otra parte, ellos, según lo afirmado por los sofistas, también serían “más racionales” que los débiles ya que, a diferencia de estos últimos, sí se han enterado del engaño que se escondería tras la moral y sabrían cómo utilizarla ideológicamente para controlar a quienes no pueden ponerse por encima de ella (pp. 315 y ss.). Esta tesis, que rechaza Gauthier, significaría que el contrato social es en realidad un falso acuerdo y, en mi opinión, nuestro autor se percató de las consecuencias catastróficas que esto traería no sólo para su propia propuesta, sino también para cualquier teoría filosófica que, en general, intente hallar un fundamento *racional* de la moralidad. De allí que reconozca que su maximizador condicional MR podría ser señalado, precisamente, como ese tipo de individuo que, al igual que los poderosos del relato sofista, no está interesado en someterse a las restricciones morales por mor de ellas mismas, sino que las asumiría como un mero mal necesario o como una fachada para ocultar sus verdaderas intenciones.

¹⁰³ Como ya se mencionó en el capítulo anterior, para lograr esta igualdad en la racionalidad, Gauthier cree que bastaría con que, o bien los agentes menos ilustrados, o las culturas menos desarrolladas tecnológicamente alcanzaran un desarrollo igual al de las culturas y los agentes más avanzados, o bien sería cuestión de corregir las falsas creencias que tengan las culturas y los agentes menos desarrollados.

De manera que al interior de su propio discurso, y si Gauthier quiere salvar su propuesta una vez se ha permitido mencionar la peligrosa tesis de la sospecha, se hace entonces urgente hacerle cambios importantes a su *modelo de agente*. Pues debe mostrar cómo es que éste sí contiene los elementos de sociabilidad que le permiten superar los cortos alcances del *homo oeconomicus* y, ya con esto, podría entonces ofrecer –más exitosamente– una explicación no-sofista o no-hobbesiana de la motivación que tiene un agente para ser moral (pp. 316 y ss.). En lo que sigue de esta exposición el lector juzgará si este nuevo intento de Gauthier prospera o no. Pero, en mi opinión, habría grandes dificultades para ello, teniendo en cuenta la fuerza que desde el principio le ha dado tanto a su idea del *homo oeconomicus* autointeresado y no preocupado por otros, como a su modelo de sociedad basada en el *mercado*. Le propongo al lector examinar si, con los ajustes que a continuación Gauthier intenta hacerle a su modelo, logra superar lo que, en mi opinión, son cinco de las grandes dificultades que trae el haberse comprometido, como lo ha hecho hasta el momento nuestro autor, con su modelo de agente económico. A saber:

- 1) El ya mencionado problema de su motivación para la moral.
- 2) El resultado escéptico de reconocer que, finalmente, no hay razones para respetar las restricciones morales si se es un *homo oeconomicus* medianamente lúcido.
- 3) El hecho de que el modelo asumido por Gauthier y que él mismo reconoce como una reedición del modelo sofista y hobbesiano del hombre “natural”, sólo incluye un repertorio muy reducido de las acciones y decisiones humanas moralmente relevantes, dejando por fuera muchas otras. Las cuales, sin embargo, creo que no sería fácil descartar ni como “irracionales”, ni como pertenecientes a esferas ajenas a la moral, pero que, de ser ignoradas, se perdería de vista buena parte de aquello que nos interesa examinar en la experiencia moral.
- 4) Esto último implica también que este tipo agente, que parece tener una vida moral bastante pobre, prácticamente no tendría recursos –conceptuales o de criterio– para llevar a cabo una *reflexión* moral propiamente dicha. Esto es, para *deliberar* y decidir cómo actuar en circunstancias moralmente relevantes que requieran del individuo mucho más que un mero cálculo de costes-beneficios, o que hagan necesaria su deliberación sobre criterios de justicia, de moralidad o de razonabilidad en aquellas situaciones en las que cada agente tenga que hacer el esfuerzo por llegar a formularse un

juicio ponderado. Si en tales situaciones se hace necesario aclarar y discutir aquello que haría justo/injusto, moralmente aceptable/inaceptable un curso de acción, una norma, una institución, o una decisión sobre bienes, valores o intereses que bien pueden entrar en conflicto o ser inconmensurables, no se ve, creo yo, qué utilidad pueda ofrecer o qué luces pueda prestar, para tal reflexión o deliberación, el escueto esquema que parece manejar el *homo oeconomicus* de Gauthier.

5) Los agentes morales que de esta manera modela el autor, agentes cuya vida moral no aparece o, si lo hace, no incluye las complejidades a las que arriba aludo, siguen careciendo de contexto y de vínculos previos; siguen siendo los Robinsones cuyos lazos se asumen como artificiales e inestables, no pudiéndose, por tanto, explicar tales lazos sino de la forma en que parecen hacerlo los sofistas del relato platónico.

Creo que sumadas estas cinco falencias del agente de Gauthier, se mostraría como un modelo poco exitoso desde el punto de vista de su *plausibilidad ética* y, por tanto, nuestro autor no obtendría el resultado que buscaba al escribir su *Morals by Agreement*. Pues, recordemos, el fin de esta obra es presentar una propuesta propiamente *ética* que esté en posición de competir seriamente con las diversas posturas filosóficas desde las cuales se ha intentado fundamentar *racionalmente* la moral. Ante las dificultades para lograr su cometido, Gauthier intentará, como última salida, que su *homo oeconomicus* se transforme en un agente que supere los problemas antes señalados. Veamos entonces, a continuación, cómo se desarrolla esta suerte de “transformación” y si con ello se solucionan dichos problemas.

2. El tercer paso en la moralización del *homo oeconomicus*: su transformación en el individuo liberal

Gauthier reconoce que tal y como la ha mostrado hasta ahora, su figura del agente moral-racional como *homo oeconomicus* es fundamentalmente la del “apropiador” (*appropriator*): un ser caracterizado por su afán de apoderarse de una cantidad siempre creciente de recursos y por la búsqueda incesante del dominio sobre éstos, en detrimento del que pudiesen tener los demás agentes. Este “apropiador”¹⁰⁴ orienta todas sus acciones a la satisfacción de sus propios deseos, cada vez más costosos

¹⁰⁴ De modo análogo a lo que me sucede con el vocablo *tuism*, me permitiré castellanizar de extraña manera el término *appropriator*, pues no quisiera perder el sentido que tiene para Gauthier.

o exigentes, pero de un modo tal que se ahorre los correspondientes costes, trasladándoselos a los demás agentes con los cuales interactúa y a quienes afecta con sus actividades.

Economic man, whether capitalist or worker, is an indefinite appropriator, seeking to subdue more and more of the world to his power exercised in the service of his preferences (p. 316).¹⁰⁵

Gauthier advierte que esta figura es sólo una entidad teórica, una idea bastante útil para explicar e ilustrar ciertos fenómenos, ciertos comportamientos y formas de interacción: no se trata de un sujeto que pueda ser localizado como un hecho real. Sin embargo, el autor también reconoce que la ficción del *homo oeconomicus* ha terminado por constituirse en una idea increíblemente ilustrativa de la manera en que nos concebimos a nosotros mismos. Ha calado tan hondo en nuestra conciencia como occidentales modernos, que ha logrado moldear buena parte de lo que hacemos o de lo que pensamos que hacemos o que hacen las demás personas, rebasando así los límites de un modelo utilizado, en principio, con fines meramente especulativos.

[. . .] Economic man has entered into the ground of our self-consciousness, of the way in which we, heirs to a tradition of social thought that has matured in Western Europe and America since the seventeenth century, understand ourselves . . . economic man is a caricature or distortion, who has come more and more to shape our reality (p. 316).

Creo que el lector estará de acuerdo en lo acertado y, por demás lúcido, que resulta este diagnóstico de Gauthier acerca del poder que ha adquirido la ficción del *homo oeconomicus*. Pero también pienso que esto mismo nos obligaría a cuestionar el hecho de que nuestro autor haya comprometido su discurso, desde su planteamiento inicial y a lo largo de nueve de los once capítulos de *La moral por acuerdo*, con una figura que a la que el propio filósofo canadiense ahora reconoce como caricaturesca, incluso a pesar de que ésta haya ejercido, repito, una innegable influencia en la

¹⁰⁵ Resulta interesante que Gauthier, acaso adelantándose a alguna suerte de crítica marxista, iguale el comportamiento, según él, autointeresado y voraz del trabajador, con el que suele atribuírsele al capitalista, por el simple hecho de que ambos serían seres económicos.

formación de la mentalidad moderna.¹⁰⁶ Cabe, pues, objetar el compromiso que tan fuertemente asumió Gauthier, tanto con esta pintura del *yo-del-mercado* a la que alababa con entusiasmo antes de arribar al décimo capítulo de su libro, como sus anteriores y vehementes apologías del contexto en el que, obviamente, el autor situaba a su *homo oeconomicus*: el *mercado*, entendido como un sistema libre de trabas políticas con las que se podría —equivocadamente, en opinión del canadiense— intentarse evitar las inequidades sociales (pp. 99-112).¹⁰⁷ Recordemos que en dichas apologías nuestro filósofo mostraba a esta forma de mercado como el ejemplo a seguir, como el ideal de interacción social no-coercitiva y, por tanto, moral y racionalmente defendible. Si a estas alturas del desarrollo de su discurso Gauthier ya cae en cuenta de las desventajas que trae, para la plausibilidad de su propuesta inicial, su modelo de *agente*, entonces también es posible que repare en la siguiente crítica. A saber: si el agente de la moral por acuerdo es básicamente el *homo oeconomicus* y si ocurriera que este personaje se viera atrapado en alguna situación que le impidiese manipular a los demás agentes con los que pacte, a fin de que éstos hagan honor a los mandatos morales, o si al *homo oeconomicus* le sucediera que, por diversas razones, no lograra aparentar convincentemente ser un maximizador restringido MR, entonces es factible pensar que haría caso omiso de la moral. Incluso hasta podría optar por imponer su voluntad mediante el uso de la fuerza. A ello le empujaría su “naturaleza” que, tal y como parece haber sido asumida por el propio Gauthier y no sólo por los sofistas, es “nuestra” naturaleza; una naturaleza que desde el principio ha socavado la posibilidad de la moral; ha hecho difícil que pueda hablarse en serio de una moral “racional”.

¹⁰⁶ En *Modern Social Imaginaries* (pp. 15 y 69-80) y en *A Secular Age* (pp. 176-185), Ch. Taylor intenta mostrar la fuerza de esta idea para modelar nuestra autoimagen como cultura, pero también como aquello que en los albores de la modernidad jugó el papel de una pieza fundamental dentro de un discurso con el que se buscaba legitimar y proponer un nuevo orden social, unas nuevas relaciones políticas y unas prácticas económicas orientadas a cambiar aquellas que eran propias del *ancient regime*. Dicho discurso pudo haber sido novedoso e inspirador en tiempos de Locke, pero hoy en día tal vez nos parezca casi “natural”, incluso si asumimos una posición crítica y pesimista frente a él. Acaso a algunos nos aterre que ciertas prácticas sociales que nos caracterizan como occidentales modernos sean tan parecidas a las de ese, en principio, modelo o caricatura que pudiera ser el *homo oeconomicus* en su sentido teórico y, sobre todo, en una versión mucho menos simpática, menos pacífica y menos constructiva que aquella que aparecería, v.g. en Locke o en Smith. Sin embargo, tal vez parte de lo que nos aterra podría deberse precisamente al hecho de que ese personaje se nos haga menos ficticio de lo que quisiéramos; a nuestro temor de que se halle mucho más cerca de haberse “concretado” en el tipo de interacciones que sostenemos en nuestro día a día o que, al menos, pensamos que sostiene “el común de la gente” en nuestras actuales sociedades capitalistas.

¹⁰⁷ Véase también: pp. 260-265, 270-285, 289-298.

En este mismo orden de ideas, nuestro autor puede verse confrontado por la objeción de que, si en algunas circunstancias las verdaderas intenciones del *homo oeconomicus* son públicamente desveladas y su “secreto” dejase de ser tal, esto es, si ya no solamente fuese conocido y utilizado a conveniencia por el astuto MI que intenta manipular a sus congéneres, con el fin de que éstos sí cumplan con su parte del contrato social, mientras él sólo finge hacerlo, entonces ya muchos (o tal vez todos) podrían hacer mofa de dicho contrato; al quedar descubierto el engaño, ya no produciría el efecto deseado por el *homo oeconomicus*. Creo que a la vista de semejante conclusión, el lector bien puede insistir en preguntarse lo que arriba ya se cuestionó: ¿qué tan acertado puede haber sido, entonces, el haber edificado la posibilidad de una moral racional sobre la base de una concepción de la naturaleza humana como algo que, de suyo, hace tan difícil concebir la moral como algo distinto de una máscara o un mero instrumento de dominación?

. . . [That] the natural response to what is claimed to be a necessary evil is to make it unnecessary. As the morality of economic man, a moral by agreement is a sham (p. 316).

Por lo anterior, repito, Gauthier tal vez se vea en la necesidad de tomar distancia de su modelo inicial de *agencia* humana, si no quiere arribar a la conclusión escéptica que ahora parece preocuparle tanto. Creo que lo relevante de este cambio al que, por lo visto, se ve forzado nuestro autor estaría en que, si en un principio manifestó que su propuesta estaba pensada exclusivamente para *maximizadores autointeresados y no-concernidos por otros* agentes, resulta que ahora, en este punto de su discurso, se produciría un giro que nos llevaría a un resultado al que acaso no pueda menos que calificarse de sorprendente: la conclusión de que aquello que garantizaría que cumplamos con la moral en general, y no sólo con la moral por acuerdo es, paradójicamente, el hecho de que *no* somos ese tipo de maximizadores. Por tanto, y en contra de lo que al inicio de su texto Gauthier enfáticamente dijo que se proponía demostrar, ahora lo que vemos es que aquello que permitiría que deseemos ser morales y esperemos que las demás personas también lo sean, no son razones atadas al autointerés y al afán maximizador propio de los egoístas yoes-del-mercado. Pues dichas razones, de suyo y tal como nos lo muestra la lección sofista, imposibilitan un

compromiso estable y sincero con la moral. Este reproche que, en mi opinión, el lector bien puede hacerle a Gauthier, daría al traste con toda su empresa de fundamentación de la moral como racionalmente justificada, en el sentido de cimentarse sobre la base de una racionalidad específicamente *maximizadora* y de no obedecer a consideraciones de carácter “tuista”.

Ante esta objeción, nuestro autor intenta hacernos ver que este –posiblemente desastroso– resultado no tiene por qué producirse, ya que para impedirlo habría que contar con algo en lo que Gauthier no había insistido anteriormente pero que, a estas alturas, sería fundamental que se demostrara de manera mucho más clara: el hecho de que nosotros, los agentes humanos, si bien somos *yoes-del-mercado*, también ocurre que *no lo somos del todo*. Por ello, el filósofo canadiense anuncia que la solución a este problema estará en que pueda mostrarse cómo es posible esta doble condición en uno y el mismo agente; dualidad que tal vez muchos de nosotros encontraría bastante plausible, sobre todo tratándose de los sujetos que viven en una sociedad en la que, como la nuestra, y como arriba se comentó, ha hecho una carrera tan exitosa la idea del *homo oeconomicus*. Sin embargo, repito, si hacemos memoria, esta dualidad *no* era propiamente la que proporcionaba, desde su mero inicio y antes del capítulo X de *La moral por acuerdo*, la posibilidad de una moral justificada racional y no teológica, ni metafísicamente. Antes, casi que al contrario, la promesa inicial de nuestro autor era que se nos iba a demostrar que la *única* vía *racional* por la cual llegamos a comprometernos a respetar códigos morales, es aquella a la que nos llevan nuestras motivaciones *autointeresadas*, *i.e.* por completo ajenas a los sistemas tradicionales o no-filosóficos de moralidad, al tiempo que para la satisfacción de dicho autointerés nos servimos *exclusivamente* de una racionalidad *maximizadora*. Lo cual, a su vez, explicaba que, antes de que se hubiera llegado a este punto de inflexión en su texto, nuestro autor se sentía plenamente autorizado para descalificar con vehemencia, tanto aquellos discursos que apelan a *otras* supuestas *racionalidades* (v.g. Parfit), como a toda propuesta ética que en la que se postulen *otras* formas de *moralidad-racional* (v.g. Kant).

Por lo anterior, ahora que Gauthier intenta dar un giro en su argumentación, de modo que logre demostrar que los agentes humanos poseen la doble condición de unos *yoes-del-mercado* que, por lo visto, ya *no lo son tanto*, el autor anuncia que hará algunos ajustes importantes en su modelo inicial del agente. Así, éste deberá incorporar,

a partir de esta última parte de su discurso, aquellos elementos de nuestra agencia racional-moral que *no* obedecerían al modelo del *homo oeconomicus* que aparecía al inicio de *La moral por acuerdo*.

Morals by agreement would afford economic man a beneficial constraint, if only he could be constrained. Because we real human beings share some of his characteristics, morals by agreement afford us a beneficial constraint, and because we are nevertheless not economic men and women, we can be constrained. Morality does not undermine itself (p. 317).

2.1 La necesidad de “pacificar” el mercado

No obstante lo anterior, pienso que un lector no muy satisfecho podría seguir insistiendo en que esta necesidad en la que ahora se ve Gauthier, necesidad de incorporar aspectos no-autointeressados y no-maximizadores en la figura del agente racional que se decide a actuar moralmente, después de todo sólo demostraría que la posibilidad de dicha decisión y actuación descansa, como ya se ha sugerido, sobre esos *otros* aspectos ajenos al egoísmo y al cálculo maximizador. Así, el hecho de que nuestro autor recurra a estas nuevas formas de motivación humana, con el fin de reajustar su modelo de agente para que éste no haga imposible cualquier sistema de moralidad racional, incluyendo a su moral por acuerdo, justamente apuntaría a demostrar la tesis *opuesta* a la que él en principio defendió. No ocurriría, entonces, que nos decidimos a ser morales –en último término– por consideraciones extra o pre-morales, sino que, por el contrario, tales consideraciones constituyen un obstáculo –incluso destructivo– para la moral; y un obstáculo al que sólo se le puede hacer frente mientras no se funde sobre él la posibilidad de una motivación para cumplir con principios morales y para esperar, por parte de las demás personas, ese mismo cumplimiento.

De este modo, solamente si tenemos abierta la posibilidad de acudir a *otros* motivos básicos que expliquen la capacidad humana para la moral y que *no* sean las razones del *homo oeconomicus*, entonces sí podrían tener solución los escollos que suponen para la moral los ímpetus del yo-del-mercado. Si sucede lo contrario, es decir, si aquello que Gauthier llama “motivaciones primarias” constituyen la base motivacional *última* de los agentes racionales, esa que, según él, resulta ineludible

porque subyace a *toda* motivación humana, entonces pienso que no puede verse a la moral como una empresa o una praxis racionalmente justificada. Podría justificarse de otras maneras, por ejemplo, como un mal necesario, tal y como se vio en el caso de los sofistas, pero no como algo a lo que deseamos, de suyo, contribuir; no como un marco normativo en el que creamos sinceramente, a menos que seamos parte del rebaño inocente que sigue sin enterarse de la mala noticia que nos cuenta Trasímaco.

Probablemente Gauthier no encontraría justificado este reproche, si bien y como se ha reiterado, a estas alturas de su discurso ya ha puesto su atención en él, y de allí su nueva estrategia de hacerle ajustes a su modelo de agente. Nuestro autor expresamente anuncia que dicha estrategia requerirá que se complementen entre sí lo que él vería como las *dos fases* de la exposición de su “nueva” figura del agente. A saber, por *un lado*, estaría una descripción más centrada en sus capacidades *sociales*, capacidades que el *homo oeconomicus* no tiene y que son las que, en cambio, sí caracterizan al tipo de agente que Gauthier llama el “individuo liberal” (pp. 317 y 345-349). Para este personaje, que será el modelo con el cual el autor anuncia que se comprometerá definitivamente, en contraste con su anterior figura del agente, la moral no será un mal necesario ni, menos aún, un perverso instrumento de dominación. De allí que, por el *otro lado*, Gauthier desarrollará la *otra face* de la exposición de su agente “renovado”: las notas fundamentales de la *sociedad* en la que éste actúa, es decir, el contexto que le caracteriza y que constituye aquello que el autor define como una “sociedad justa”. Se trata de una sociedad en la cual deben verse favorecidos, tanto el desarrollo de un *ethos* individualista, como la libre acción de cada individuo y, en general, el despliegue efectivo de las diversas formas de libertad personal. A continuación se examinará de qué manera las características más importantes de esta sociedad establecida por los agentes liberales, se entrelazan necesariamente con las notas fundamentales que distinguen a dichos agentes.

Como ya se ha anunciado, nuestro autor intentará desarrollar lo que creo admite ser visto como una suerte de *transición* entre la figura hasta ahora presentada de su *homo oeconomicus* y su nueva versión –acaso la definitiva–, de su individuo liberal, el cual será, entonces, algo así como el producto de una serie de transformaciones aplicadas a su inicial modelo de agente. La exposición de esta “metamorfosis” debe mostrar de manera clara, según Gauthier, que su propia y “mejorada” versión del *homo*

oeconomicus permite que este personaje no sea en absoluto asimilable ni al hombre “natural” del que hablaban los sofistas, ni al agente en el que pensó Hobbes (pp. 317-321). Sin embargo, en mi opinión resulta un tanto sorprendente la explicación que de ello ofrece Gauthier, dados los problemas en los que se vio al suscribir un modelo de la agencia humana tan atado al modelo del mercado.¹⁰⁸ Digo que puede resultar sorprendente este viraje en la argumentación del autor canadiense, pues éste sostiene que, a diferencia de la figura que aparece en el discurso de los sofistas, y dadas las

¹⁰⁸ Al respecto, habría que hacer una salvedad que espero se vea mejor justificada en lo que sigue de este capítulo. Creo que se puede advertir una cierta tensión entre dos nociones de “mercado” que suelen alternarse a lo largo del texto de Gauthier. A saber: 1) Una idea del mercado entendido como fundamentalmente competitivo, dada la escasez de recursos y la psicología esencialmente egoísta de los agentes que interactúan en dicho escenario, el cual se constituiría, por tanto, en un juego de interacción del que *no* todos salen beneficiados. Esto haría que el mercado difícilmente pudiera ser equiparable al Estado moderno que, para autores como Hobbes, pone fin al estado de naturaleza. Este último, por tanto, sería más cercano a la competitividad del mercado y al egoísmo de sus agentes, caracterizándose así por los riesgos que se busca conjurar mediante el contrato social y el Estado que se legitima gracias a éste. Y, 2) Habría otra versión del mercado, más cercana a Smith o al mismo Locke, y en virtud de la cual el intercambio económico —en general— es visto como una parte fundamental de *toda* sociedad armónica y pacífica. En este último caso, y como ya se vio (en la sección 3 del capítulo 1, y la sección 1 del capítulo 2), Gauthier parece equiparar el mercado al Estado que surge luego de que se pacta el contrato social y que pone fin a la situación de riesgo generalizado, propia de un previo estado de naturaleza, tal y como parece ser pensado por Hobbes. En consonancia con esto, y como de alguna forma ya se mencionó, a pesar de que el filósofo canadiense toma prestada la noción lockeana del *proviso*, empero no menciona otro aporte del autor del *Tratado sobre el gobierno civil*: la idea de una sociedad “originaria” no-violenta, cuyos integrantes, al contrario de lo que pienso que puede verse en *La moral por acuerdo*, gozan de unas libertades y unos derechos también “originales” que, en un sentido normativo, *deberían* seguir siendo protegidos por el Estado moderno que nazca posteriormente gracias al pacto social. Recordemos que, para Locke, incluso dicho Estado ganaría su base de legitimidad si cumple con el cometido de salvaguardar esa igualdad y esas libertades “originales” que son vistas, entonces, como propias de un estado de naturaleza modelado de una manera distinta a la que parece perfilarse, tanto en el discurso de Gauthier, como en su versión de Hobbes y de los sofistas. Así, por contra de lo que puede apreciarse en estos pensadores, en cuanto a sus ideas sobre el origen de toda asociación humana (de la sociabilidad humana en general), desde el punto de vista de Locke es precisamente en virtud de dicha asociación humana “original” y pacífica, asociación cuyas bondades debe seguir protegiendo el Estado que se funde posteriormente gracias al contrato social, que a todos nos son debidos una serie de derechos sin los cuales es imposible garantizar el bienestar de todos y de cada uno. Estos derechos, que son *previos* a dicho Estado, sin embargo, repito, son la base de legitimidad de éste, hasta el punto en que se puede apelar a ellos cuando tenga que justificarse una rebelión en contra de aquellos mandatarios que no respeten esos derechos “originarios”. En Gauthier, insisto, a mi entender, y a pesar de su apropiación del *proviso* lockeano, no se le da juego a esta idea de un estado de naturaleza no-violento en el que, por tanto, ya nos son debidos unos derechos “originarios”. Y ello, aun cuando el filósofo canadiense algunas veces identifique el estado de naturaleza con el mercado, cuando se trata de una de las dos nociones de mercado a las que, de tanto en tanto, apela: la más halagüeña, la más pacífica o armónica de ambas. Por lo anterior, aunque nuestro autor no la mencione, la idea del intercambio económico como parte de la armonía que puede haber entre unos agentes pensados *à la* Locke y que, a diferencia del panorama pintado por Gauthier, *no* son concebidos como *originalmente egoístas ni violentos*, se me ocurre que tal vez le hubiese podido ser de utilidad al autor canadiense, como un posible apoyo que le facilitara su nuevo cometido de “civilizar” al *homo oeconomicus*, a fin de que deje de ser un obstáculo para justificar racionalmente la moral. Sobre las diferentes concepciones del mercado tanto en Hobbes como en Locke, véase: Macpherson (1962, pp. 46-70; 194-221).

características propias del tipo de *sociedad* premoderna en la que estos pensadores vivían, el moderno *homo oeconomicus* posee la ventaja de que opera dentro del contexto del *mercado*. Lo cual marcaría, según nuestro autor, una distancia abismal entre su propia versión del agente y la que puede obtenerse del discurso de Trasímaco y los suyos.

Para Gauthier, el mercado posee una normatividad y permite unas posibilidades de ejercicio de la agencia humana que van mucho más allá del mal necesario que los sofistas veían en la moral y en las normas de la justicia. Y frente a la versión hobbesiana, el filósofo canadiense afirma que su *homo oeconomicus* tiene una vida social mucho más rica que aquella que se alcanza mediante el pacto de no agresión que pone fin al estado de naturaleza descrito por Hobbes. Gauthier cree que en su versión del contrato social y de los agentes que lo suscriben, a diferencia de lo que ocurre con el relato del autor del *Leviatán*, los agentes no son descritos como completamente “económicos”; ya que la adhesión de éstos al pacto no se debe únicamente a la búsqueda de su propio beneficio, sino también que ven en la moralidad y en las reglas de la justa convivencia una forma de atractiva sociabilidad y de expansión de los propios talentos; lo cual está lejos de asimilarse al mal necesario o la restricción indeseable, pero inevitable, que significaban para el hombre “natural” de Hobbes. Gauthier cree que su *homo oeconomicus* debe ser pensado como un individuo mucho más *sociable* que sus antecesores sofista y moderno. Se trata de una figura “impura”, más real, menos brutal. En *primer* término porque si bien no deja de ser “económico” y, por tanto, también está interesado en competir con sus congéneres, en poseer y en dominar más recursos que ellos, y en “ganar” dicha competencia, sin embargo Gauthier no cree que esto tenga que verse como una actitud *asocial*.¹⁰⁹ En *segundo* lugar, aunque es cierto que su *homo oeconomicus* también es un “apropiador”, para el filósofo

¹⁰⁹ Este carácter *social* de la competencia entre los hombres es incluso reconocido, dice Gauthier, por el mismo Rousseau (Cf. Rousseau, Trad. 2001, pp. 148-149, y la nota cero de la p. 235), quien llega a afirmar que competimos “por amor propio” no siendo por ello unos seres asociales. Por el contrario: dicha competencia es fundamental para la idea que el agente tenga de sí mismo, hasta el punto que le moldea como individuo. Uno se perfila a sí mismo, y en ello el canadiense dice suscribir una tesis del ginebrino, en su relación con esos otros con los cuales se encuentra y compete. “. . . Rousseau argues such a person must not be identified with the natural asocial man, for she exists only in her relation to others, taking indeed her very sense of self from that relation” (p. 318). (Cf. Rousseau, Trad. 2001, p. 203). A los ojos de un lector que haya seguido hasta acá *La moral por acuerdo*, creo que ahora esta súbita referencia de Gauthier a Rousseau resulta un tanto difícil de encajar con los primeros nueve capítulos del texto de nuestro autor. Al menos, pienso que se requerirían ciertas explicaciones que la hicieran ajustarse un poco mejor a lo dicho en esa primera parte de su libro.

canadiense las actividades que caracterizan a dicho personaje no son necesariamente asociales, dado que éstas se desarrollan en un contexto muy especial: el *mercado*. De nuevo, Gauthier nos previene de no asustarnos: no debemos temer que resurjan los ya mencionados problemas que acarreaba el modelo del yo-del-mercado, pues, según el autor, dicho sistema funciona gracias a la figura de los *contratos*. Lo cual permite que el moderno sistema capitalista se constituya en un marco de acción bastante distinto del estado de naturaleza en el que los agentes involucrados buscaban, mediante la fuerza o el medio que estuviese a su alcance, apropiarse de todos los recursos que pudieran (incluyendo a su prójimo).

Para el filósofo canadiense, si bien el mercado no es una comunidad de ángeles o de seres dedicados a hacer felices a sus semejantes, sí opera sobre la base de unas reglas que restringen la conducta maximizadora de quienes en él participan. Esto, por supuesto, es cierto que requiere de suyo una mutua instrumentalización entre los agentes, pero no hay que olvidar que se trata una instrumentalización regulada y recíproca, de manera que las relaciones así establecidas —que son relaciones contractuales— no van a ser asimétricas ni coercitivas (pp. 321-324). De esta forma, nos hallaríamos bastante lejos, afirma nuestro autor, de la mera apropiación llevada a cabo por el hombre “natural” de los relatos sofista y hobbesiano. En conclusión, y aunque Gauthier no lo anuncie en estos términos, podemos ver que suavizar su figura del *agente*, esa que le ha acarreado tantos problemas, requeriría de un paso importante: hacer menos desagradable su pintura del *mercado*. Que este último no sea asimilable al violento estado de naturaleza permite, según el canadiense, que la competencia que surja como parte esencial de la dinámica del mercado no sea una competencia a muerte o entre enemigos que busquen anularse radicalmente. De hecho, no es una competencia ni entre enemigos ni entre amigos sino, simplemente, entre seres que buscan su propio beneficio y no están ocupados en atender al de los demás. La única “preocupación” mutua de la que cabría hablar aquí sería aquella que se expresa en los contratos. Esto permite que se presente otra gran diferencia, en opinión de Gauthier, entre su *homo oeconomicus* y sus antecesores sofista y hobbesiano; a saber, que el primero no compite en juegos de suma cero. No ocurre que si “gana”, entonces los demás agentes “pierdan” y se queden por fuera del juego en cuestión, puesto que no se trata de juegos completamente competitivos. Esta circunstancia se implica mutuamente con la

percepción que el *homo oeconomicus* de *La moral por acuerdo* tiene de sus nexos sociales, ya que este personaje ve a la sociedad como una red de relaciones contractuales, de mutua *instrumentalización*. Se trata de relaciones que por su misma definición y puesto que están reguladas por *contratos*, de *un lado*, se eligen *libremente*, *sin coerción* y, del *otro*, implican que quienes así las eligen *no estén concernidos los unos por los otros*.

Aquí de nuevo nuestro autor insiste en su idea de que este esquema de lazos sociales es uno de los inventos más liberadores que ha engendrado la humanidad ya que, gracias a él, los nexos que se establecen entre los agentes así relacionados —económica o contractualmente— no dependerán de sus vínculos de afecto, de clan o de tribu, sino sólo de su voluntad individual y de lo que consideren conveniente para sus intereses meramente personales. Ahora bien, Gauthier reconoce que este favorable aspecto libertario, esta emancipación de los lazos tribales también tiene su revés y su peligro: la eventual incapacidad que pueda tener el yo-del-mercado para aquellas relaciones que no obedezcan a la lógica económica, incluyendo las relaciones de afecto que, en principio, serían objeto de una elección voluntaria (p. 326-329). No obstante esto último, el balance que obtiene el filósofo canadiense es claramente positivo: habría que concluir, según él, que las ventajas que ofrece su nueva versión del *homo oeconomicus* harían que éste aparezca como una figura más rica, compleja y plausible que sus antecesoras.

Para explicar esta afirmación, de nuevo nuestro autor apela al *mercado*: si éste es el contexto de interacción en el operan los *agentes* que Gauthier ahora intenta dibujar siguiendo su nueva versión del *homo oeconomicus*, entonces las ventajas que el canadiense atribuye a dicho mercado hacen que quienes allí cooperen no puedan ser equiparables a los agentes en los que pensaban los sofistas o Hobbes. Para el autor de *La moral por acuerdo*, esto se debe a la determinante y positiva influencia que el mercado ejerce sobre la *capacidad de agencia* de quienes actúan dentro de él. Según Gauthier, si se examinan los discursos de algunos pensadores modernos, como Rousseau y Hobbes, puede verse que éstos no cayeron en cuenta de la amplitud de posibilidades que se abren ante el *homo oeconomicus* gracias a la riqueza y complejidad de la sociedad de mercado (pp. 319-320). El filósofo canadiense afirma que el ginebrino falla al exagerar las consecuencias de las *desigualdades* sociales, sin reparar en que aquellos agentes que sean menos poderosos o resulten menos favorecidos también

pueden salir beneficiados de su participación en un tipo de sociedad fundamentada sobre la base de un contrato social; sociedad que se encuentra mucho más adelantada que sus antecesoras.

Por otro lado, Hobbes fallaría, en opinión de Gauthier, porque no es claro en mostrar que aquellos agentes que en el estado de naturaleza podían apropiarse de más recursos y tener menos limitaciones de las que tienen luego de que se firma el pacto social, viéndose con ello obligados a renunciar a buena parte de su antiguo poder, es cierto que pagan con dicha renuncia el precio de la paz; pero también es innegable que, a cambio de eso gozan, precisamente, de las innumerables y evidentes ventajas de una sociedad pacífica, estable y más segura.

Each person may reasonably expect to do better as a consequence of the benefits of market and cooperative interaction, even if these benefits are not fairly distributed. If among economic men moral and political constraints come to serve as instruments of domination, perpetuating and strengthening inequalities, yet they may also serve as instruments of improvement (p. 320).

Nuestro autor no se detiene a mostrar qué tipo de ventajas, “mejoramiento” o “progreso” (*improvement*) representa la sociedad de mercado para quienes resulten *menos favorecidos* en ella y, en este sentido, creo que parece dejar sin respuesta las objeciones que le haría una persona preocupada por las mismas razones que preocuparían v.g. a Rousseau o Marx. A cambio de ello, Gauthier insiste en atribuirle al mercado una innegable cualidad: un progreso constante en virtud del cual dicho sistema se constituye en una compleja, amplia, cambiante e inagotable fuente de recursos o de bienes. Con ello, y en franco contraste con el estado de naturaleza tal y como lo pensaban los sofistas y Hobbes, el mercado no resulta ser un medio cuyos recursos se agoten, ni tampoco sería un mero objeto de depredación por parte de los hombres que en él compiten. Para el filósofo canadiense, a los ojos del autor del *Leviatán*, así como para los contradictores de Sócrates, la paz es la *única* ventaja que le ofrece al hombre natural la sociedad que sustituye al estado de naturaleza. Empero, Gauthier cree que esa paz que sigue al contrato social también va acompañada del goce de *otras* ventajas adicionales, pues la sociedad de mercado ofrece una gran sofisticación de los modos de producción, una complejización de las actividades humanas y el constante surgimiento de nuevos recursos y de todo tipo de bienes —no sólo materiales—, con lo cual no ocurre que el

medio en el que conviven los hombres se vacíe o se deprede. Tampoco ocurre que los agentes más afortunados acaparen y agoten todos los bienes, mientras que los que cuenten con menos suerte resulten privados de absolutamente todo lo que necesiten. En un medio así, según nuestro autor, las mayores posibilidades de expansión de los talentos y de las actividades productivas y creativas de las personas permiten que, además de la paz, haya una amalgama enorme de ventajas frente a la sociedad en la que pensaban los sofistas y Hobbes. De manera que, aunque en el mercado los bienes también son escasos y no está asegurada una igualdad en la distribución de éstos, es innegable que allí se da un hecho que marca una enorme diferencia frente al estado de naturaleza: una infinidad de actividades cooperativas y de interacciones humanas que son bien distintas a simples juegos de suma cero.

2.2 El aspecto institucional, la persistencia del egoísmo y las desigualdades sociales: ¿la autonomía en riesgo?

Gauthier afirma que las mencionadas actividades e interacciones, junto con las instituciones que las organizan y regulan, hacen que, por una parte, la producción de bienes –y, repito, no sólo los materiales– sea siempre objeto de mejoramiento, no de agotamiento y, por la otra, que quienes participan en las diversas interacciones no “pierdan”, si bien estarían compitiendo con los demás agentes y pueda darse, por tanto, la impresión –falsa, según el canadiense– de que haya favorecidos y desfavorecidos, ganadores y perdedores. Pues nuestro autor piensa que, como ya lo hemos mencionado, las desigualdades que se producen también van acompañadas de las ventajas de vivir en una sociedad mejor y con permanentes posibilidades de progreso. De allí también que, según él, se vea con claridad que para su *homo oeconomicus*, a diferencia del hombre natural de los sofistas y de Hobbes, la vida en sociedad y las restricciones impuestas a la búsqueda de su propio beneficio no son simplemente un mal necesario, sino parte de los arreglos sociales que le permiten tener una vida mejor y más rica. Dejo simplemente anotado un comentario sobre un asunto que seguramente puede preocupar al lector que se encuentre con esta entusiasta defensa de la sociedad de mercado. A saber, la ya mencionada insensibilidad de nuestro autor frente al problema de las *desigualdades* sociales y su posible ceguera ante las consecuencias que dichas desigualdades traerían

en contra de algo que, paradójicamente, Gauthier dice considerar de la mayor importancia: la *libertad* de los individuos y el despliegue de su *capacidad de agencia*.

El autor no parece tener en cuenta la siguiente objeción que seguramente el lector habrá anticipado: si algunos agentes están en desventaja frente a otros, podría no ser tan fácil considerar que los primeros son realmente “libres” o no coaccionados/coaccionables por parte de los más aventajados. Por tanto, tal vez resultaría difícil seguir sosteniendo que los menos favorecidos pueden desplegar *sensu stricto* su capacidad de agencia, al menos no en la misma medida que pueden hacerlo aquellos que tienen mejor suerte que ellos, razón por la cual estos últimos podrían aprovechar su posición de ventaja para explotar y/o someter a los menos afortunados. Por otra parte, la historia nos ha mostrado que las fuentes de recursos para la producción de bienes o para la satisfacción de necesidades, bien pueden ser objeto de depredación y agotamiento, al tiempo que estos excesos en su aprovechamiento no siempre redundan en una distribución racional de dichos bienes sino que, más bien al contrario, el acaparamiento y la sobreexplotación de recursos pueden significar la pauperización extrema tanto de los hábitats, como de los grupos humanos; y estos fenómenos no necesariamente se le pueden achacar a condiciones premodernas o precapitalistas.

Como digo, Gauthier no hace ninguna mención de estos posibles problemas, si bien reconoce que aún sigue sin responder a una objeción que ya ha sido mencionada: el peligro de que la moral propia de una sociedad de mercado pueda ser un *instrumento de dominación ideológica*. En este punto (pp. 321 y ss.) el filósofo canadiense intenta responder a Mary Gibson (1977), quien afirma, según él, que el *homo oeconomicus* no solamente intenta aumentar su propia racionalidad, perfeccionando sus métodos para la búsqueda y el crecimiento de sus propios beneficios. También se propone disminuir las capacidades racionales de sus congéneres, mediante el desarrollo de estrategias dirigidas a manipularles ideológicamente, de manera tal que las preferencias que tengan las demás personas se vayan moldeando a la medida de los intereses del astuto *homo oeconomicus*. Gauthier le replica a Gibson que ello equivaldría a suponer que todos los juegos que tengan lugar en la sociedad de mercado serían juegos de suma cero, lo cual, como arriba se comentó, él no cree que sea cierto. Además de esto, nuestro autor arguye que la sofisticación de los juegos de cooperación en los que participa el yo-del-mercado hace necesario que éste busque interactuar con agentes tan racionales como él, agentes

que posean unos hábitos parecidos a los suyos. Por ejemplo, que sean individuos disciplinados y con conciencia de sus yoes-futuros, de modo que también sean sujetos previsivos que planeen a largo plazo sus acciones pues, de no hacerlo, pondrían en peligro los planes del *homo oeconomicus*.¹¹⁰

No obstante, nuestro autor concede que, de todos modos, el interés que pueda tener el yo-del-mercado en la racionalidad de los demás agentes es completamente *instrumental*, lo cual llevaría a la situación que Gibson denuncia: que el cuestionado personaje que presenta Gauthier no tendría ningún interés en que sus congéneres sean individuos auto-reflexivos. De allí que tampoco le interese que las preferencias que éstos tengan se correspondan con aquello a lo que nuestro autor se ha referido como “preferencias consideradas”. Incluso, la preocupación que el *homo oeconomicus* tenga en las preferencias de otros agentes estaría, por el contrario, orientada a manipularlas para que respondan a sus intereses, no a los fines de sus conciudadanos (p. 322). Creo que, como puede ver el lector, seguirían en pie no sólo las objeciones de Gibson, sino también las sospechas de los sofistas, de Rousseau y de algunos más. Si alguien quisiera defender el punto de vista de Gauthier, pienso que le parecería preocupante que estas denuncias sigan sin ser respondidas satisfactoriamente, ya que con ello se desautorizaría, como ya se ha dicho, una tesis central del autor canadiense. Pues si éste insiste en que el cumplimiento de reglas morales obedece, finalmente, a una lógica maximizadora, entonces dicho cumplimiento también podría hacer parte de una manipulación ideológica o de una estrategia de dominación, con lo cual el nexo entre moral y racionalidad se vería seriamente debilitado. Ante una objeción de este tenor, Gauthier advierte que, si bien su *homo oeconomicus* posee todavía algunas de las características asociales de sus antecesores sofista y hobbesiano, no hay que olvidar que el *mundo* en el cual desarrolla sus actividades es tan cambiante y complejo que las *preferencias* de los agentes que allí participan no están dadas del todo ni permanecen siempre iguales.

Un factor fundamental para la transformación de estas preferencias viene proporcionado por la enorme influencia ejercida por las *instituciones*, las cuales pueden

¹¹⁰ Según Gauthier (p. 322), la irracionalidad de sus congéneres sería un serio obstáculo para que su *homo oeconomicus* pudiera jugar muchos de los juegos de cooperación que hacen parte de la sociedad de mercado, incluso hasta el punto de hacerlos inviables, por la carga de incertidumbre que estas interacciones conllevarían.

cambiar y moldear las preferencias. Esto implica que, dentro de estas últimas, aquellas que admitan ser vistas como las más asociales y no-tuistas, las más características del *homo oeconomicus*, son preferencias que también pueden transformarse en virtud del efecto educativo de las instituciones de una sociedad liberal. Esta circunstancia viene, además, aunada a la complejidad y sofisticación de la producción de bienes, que hace que el mercado esté en proceso de constante evolución. Así, en este medio cambiante, los agentes igualmente cambian en cuanto a sus talentos, oportunidades, gustos, dotación (*endowments*) y preferencias. Si recordamos la lección que intentó darnos Gauthier acerca del punto de vista arquimédico en el que se eligen los principios de justicia y el diseño institucional de la sociedad, entonces habría que agregar, dice nuestro autor, la conclusión de que los agentes racionales tendrían que estar interesados en que las instituciones sirvan, precisamente, a esos principios de justicia –mutuo beneficio y libertad–. De tal manera que dichas instituciones, a su vez, moldeen las preferencias de los individuos para que éstos sigan perpetuando las instituciones, con el fin de que éstas no contraríen los principios de justicia que las fundamentan.

Las instituciones deben llevar a cabo su labor de educación de un modo tal que los ciudadanos las vean a ellas como el medio imprescindible para sus fines más importantes y, por esto mismo, no quieran cambiarlas radical ni revolucionariamente, sino que simplemente busquen hacerles, de una manera gradual, los ajustes que se vayan considerando convenientes a lo largo del tiempo, con el propósito de que las instituciones sigan protegiendo los principios normativos del *proviso* lockeano. Se establece, entonces, un círculo entre unos ciudadanos que quieren perpetuar las instituciones, puesto que las ven como el medio imprescindible para la satisfacción de sus intereses, y unas instituciones que, a su vez, moldean a los ciudadanos para que éstos tengan las preferencias que tienen y perpetúen las instituciones que determinan la praxis social. Los primeros creen que las instituciones les obedecen y por ello las defienden, y éstas, a su vez, les educan a ellos para que las defiendan.

En este punto (p. 323) Gauthier cae en cuenta de la crítica que se le avecina: su modelo de sociedad, ¿sería, entonces, algo parecido al *Mundo feliz* de Huxley¹¹¹ o a una suerte de estado totalitario que contradiría por completo su idea de la sociedad liberal y del individualismo que ésta debería defender? En ese caso, ¿nos encontraríamos, pues,

¹¹¹ Huxley (Trad. 1989).

ante el escenario opuesto al de una sociedad hecha a la medida de los intereses de los individuos que la eligen como el medio idóneo para sus fines personales? Y, siguiendo esta vía, ¿con ello no sólo se estaría malogrando el individualismo que decía proponer nuestro autor, sino también, de paso, la racionalidad de sus agentes individuales, que estarían lejos de ser reflexivos, autónomos y tener preferencias “consideradas”? Ante estas dudas, Gauthier se apresura a responder que su modelo de sociedad no implica que se elimine la capacidad reflexiva de los agentes, entre otras cosas, porque dicha capacidad, según el autor, nunca puede ser eliminada del todo (p. 324). Las preferencias moldeadas socialmente nunca van a ser del todo consistentes o conciliables con las preferencias satisfechas por la sociedad: siempre habrá una brecha entre ambos grupos de preferencias. Su figura del yo-del-mercado permanecería, por tanto, para Gauthier, inmune a las críticas, puesto que, repito, a este tipo de agente le interesa instrumentalizar a la sociedad para que ésta le ayude a lograr sus fines. No obstante, el autor concede que ello no implica que su agente esté interesado en que sus congéneres sean individuos reflexivos o críticos; le bastaría e incluso preferiría que sólo fuesen *instrumentalmente* racionales. Del mismo modo, también está interesado en que las preferencias de los demás agentes no entren en conflicto con sus propias preferencias o no sean un obstáculo para la satisfacción de estas últimas.

2.3 Racionalidad “meramente instrumental” vs. racionalidad “completa”; sociedad “artificialmente” justa vs. sociedad “esencialmente” justa

Llegados a este polémico asunto, para su solución nuestro autor introduce una nueva pareja de díadas dicotómicas. Para él, puede hablarse, por *un lado*, de una racionalidad que Gauthier considera es *sólo* instrumental y que caracterizaría a los agentes de una sociedad a la que califica de “artificialmente” justa. De *otro lado*, habría una racionalidad a la que se refiere como “completa” (*full*), y que sería propia de los agentes que conforman una sociedad que el canadiense llama “esencialmente” justa.

The justice [the economic man] he pursues is artificial, the fit of persons to society, and in his pursuit he is rightly seen to threaten the full rationality of his fellows (p. 324).

Como puede advertir el lector, esta nueva pareja de dicotomías puede aparecer un tanto insólita, si se tiene en cuenta las dificultades que podrían surgir al intentarse conciliar, no sólo los polos opuestos de cada dicotomía en cuestión, sino también a estas

mismas dicotomías con dos de las tesis en las que tanto había insistido antes Gauthier. Me refiero, en *primer* lugar, a su tesis —esgrimida a propósito de su crítica a Parfit— de que *no* es válido hablar de *otra* forma de racionalidad que no sea la instrumental-maximizadora o que fuese más compleja que ésta. En *segundo* término, le recuerdo al lector otra de las tesis anteriormente enfatizadas por Gauthier, y que sería conexas con la que acaba de comentarse: lo afirmado por nuestro autor en contra del escepticismo del sofista y en contra de la moral pre-filosófica. Más exactamente, su idea de que, si bien la justicia no es un mal necesario, como sostenían Trasímaco y los suyos, tampoco ha de ser pensada como algo que los agentes racionales busquen por mor de valores sustantivos altruistas o tuistas, propios de las morales tradicionales. Pienso que ambas tesis no resultan fácilmente conciliables con esta nueva idea, ahora introducida súbitamente por Gauthier, de que podemos distinguir entre una justicia *esencial* y otra forma de justicia, la *mera* justicia *artificial*; siendo la primera aquella que trata de ser aplicada por quienes se apoyen en una forma de racionalidad más “completa” y reduciéndose, por tanto, la segunda forma de justicia, a aquella otra que intenta desarrollar quien haga uso de la segunda y más miope forma de racionalidad. Esto es, una racionalidad que sólo se orientaría a establecer las reglas del juego que rigen las interacciones entre unos agentes a los que habría que considerar como “más básicos”: los maximizadores yoes-del-mercado, que no poseen aún la sofisticación del individuo liberal. Posiblemente, Gauthier quiera hacer eco de la idea humeana de la justicia como una virtud “artificial” que únicamente se justifica dadas nuestras humanas condiciones de escasez y de tendencia al egoísmo¹¹² y, con ello, allanar el camino para hacer más plausibles los cambios que ha de introducir el autor canadiense en su figura del *homo oeconomicus*, con miras a su transformación en el individuo liberal. Pues, como se verá más adelante, en este último se apreciarán también ciertas “dicotomías” o tendencias opuestas a las que, sin embargo, Gauthier considera como compatibles y como haciendo parte del mismo sujeto; a saber: una tendencia hacia la maximización y otra hacia la sociabilidad.

De modo que con esta nueva pareja de díadas dicotómicas, referidas a formas de racionalidad y de justicia, acaso el autor intente ahora mostrar una faceta más amable, tanto de su renovada figura del *homo oeconomicus* camino a convertirse en un individuo

¹¹² Cf. Hume, David, *Investigación sobre los principios de la moral*, sección 3, parte I, §§ 13-19.

liberal, como del contexto mismo en el que este personaje actúa y, por tanto, de la normas que gobernarían las interacciones que tienen lugar dentro de dicho contexto. Con estos ajustes introducidos a su modelo de agente, Gauthier a lo mejor busca suavizar ante sus eventuales críticos la dura imagen que pueda haberles ofrecido el *homo oeconomicus*. No obstante, una vez más reitera que a esta figura se la debe seguir pensando como “no-tuista”. De allí que, por ejemplo, si bien este “nuevo” *homo oeconomicus* disfruta de los placeres propios de la vida compartida con otros agentes, placeres en relación con los cuales parecería que *no* cabe hablar de una *instrumentalización* o de una racionalidad meramente instrumental, v.g. el placer de una conversación agradable, también es cierto, insiste Gauthier, que cada agente disfruta de dicho tipo de actividades sociales porque de ellas obtiene un placer que es *individual* (p. 324). Lo cual ocurre, afirma el autor, sin que el nuevo *homo oeconomicus* vea en esas actividades sociales, ni en los placeres que de ellas se obtienen, prueba de una menesterosidad humana a la que haya que considerar, ni mucho menos, indeseable. En general, no vería como algo negativo, ni detestable de suyo, nuestra natural falta de autosuficiencia de la que, supuestamente, estos placeres sociales serían una prueba.

Creo que tal vez en este punto Gauthier esté adelantándose a críticas provenientes de fuentes –que por lo pronto y tal vez apresuradamente me permito llamar “aristotélicas”– las cuales le señalarían lo pobre, unilateral e irreal que puede resultar un personaje al que se piensa como tomando parte únicamente en interacciones económicas o contractualmente mediadas o que, de participar en otro tipo de actividades sociales, lo haría llevado por una racionalidad meramente instrumental; esto es, un ser por completo ajeno a aquellas actividades que se buscan y valoren *por ellas mismas* (y no por lo que de ellas pueda obtenerse). A lo mejor nuestro autor cayera en cuenta de que este tipo de críticas pueden desvirtuar dos de sus compromisos más importantes, los cuales estarían necesariamente conectados entre sí. A saber, y como ya se comentó, su compromiso con la idea de que la racionalidad humana es básicamente *instrumental*, y su compromiso con el modelo de agente como *no-concernido* por otros. Pienso que buena parte de las críticas a estas dos tesis de Gauthier apuntarían, precisamente, en dirección a esas *actividades participativas* y, en general mostrarían, como su *quid facti*, esos bienes que no pocas personas parecen valorar *por sí mismos y no instrumentalmente*. Por ejemplo, las conversaciones agradables, así como todas aquellas

actividades cooperativas/participativas, bienes no-económicos y formas de interacción en las que los agentes humanos parecen interesarse sin que a ello les empuje un interés instrumental ni en su prójimo, ni en las actividades, bienes e interacciones de cada caso. Mejor aún, se trataría de bienes, actividades e interacciones que demostrarían un legítimo interés que sería ajeno a lo meramente instrumental, y que contradirían la idea de un agente maximizador no-concernido por otros.¹¹³

2.4 El valor “adaptativo” e instrumental de los sentimientos morales y del gusto por la socialización ¿Un posible sesgo reduccionista y elitista en Gauthier?

Tal vez previendo objeciones como las que acaban de mencionarse, nuestro autor ensaya una estrategia que, a mi parecer, resulta bastante ingeniosa y que de alguna manera ya se había mencionado anteriormente: mostrar que las actividades participativas y el placer que las acompaña, así como la valoración positiva y aparentemente no-instrumental que se hace de ellas, no se deben a una supuesta racionalidad no-instrumental o supra-instrumental que pudiera ser considerada como expresiva de nuestra particular agencia humana. Por el contrario, y aunque a algunos les pueda parecer paradójico, dicho placer y dicha valoración tienen su explicación *última* en nuestra innegable y esencial racionalidad *instrumental*. Esto se debe a razones *evolucionarias* pues, según Gauthier, los seres humanos, dada nuestra natural menesterosidad, hemos llegado a concederle un gran valor –un valor *intrínseco*– a las actividades participativas, ya que ellas han sido *instrumentalmente* útiles, a lo largo de nuestra evolución como especie, para paliar nuestra condición de no-autosuficiencia.

Reasoning from an evolutionary perspective, we may suppose that the value we find in participation is itself instrumentally related to our insufficiency. [...] There is nothing contradictory or paradoxical, or even surprising, in an instrumental explanation of certain of our intrinsic values. Not being sufficient, then we value participation with our fellows, but we value it for its own sake and not merely as means to overcoming that insufficiency. In understanding this, we should realize how alien economic man is from our real selves (p. 325).

De este modo, uno de los ajustes más importantes que el autor reconoce debe hacerse a su inicial modelo del *homo oeconomicus* es atribuirle algo que el hombre

¹¹³ Véase, por ejemplo, las críticas de Sen (1992 y 1997), y Nozick (1995, capítulos II y V).

“natural” o asocial no tiene: el hecho de que hace una valoración *intrínseca* de sus actividades participativas. Sin embargo, creo que resulta llamativo que dicha valoración se haga, según Gauthier, y aunque a algunos les parezca paradójico, por razones *instrumentales*. A los ojos del filósofo canadiense, no debemos tomarnos tan en serio nuestra eventual impresión de que estamos ante una paradoja, ya que pronto hemos de caer en cuenta del *origen evolucionario* de esa valoración intrínseca que los agentes humanos hacemos de aquellos lazos y actividades sociales o participativas que compartimos con nuestros congéneres. Dicho origen es lo que explica que pueda afirmarse, sin contradicción alguna, que hacemos tal valoración, dado que se trata de un hecho rastreable y producido sin que el propio agente “mejorado” de Gauthier se dé cuenta de ello. Es algo que ocurre, pues, a espaldas suyas, inconscientemente, y aún a pesar de que el agente mismo lo contradiga verbal y sinceramente en los juicios de valor que emite en su día a día; juicios en los que seguramente no le daría la razón a nuestro autor, manifestando de manera expresa que valora intrínseca y no instrumentalmente la participación con sus congéneres.

Pienso que en este punto uno podría decir que, aunque Gauthier no lo afirme de forma explícita, dicha valoración tiene dos explicaciones: una a la que podríamos referirnos como “superficial” y otra que podemos llamar “profunda”, siendo esta última la “verdadera” explicación. Creo que el intento de plantear este esquema de una *doble explicación* obedecería al ya comentado aspecto *reduccionista* del proyecto de nuestro autor. Siguiendo dicho esquema, la primera forma de explicación, la “superficial”, vendría dada desde una forma de *racionalidad no-instrumental* y desde unos *valores intrínsecos*, los cuales se corresponderían con el punto de vista del *agente* (ahora “renovado”) de Gauthier. En cambio, desde la segunda forma de explicación, que sería aquella que se correspondería con el punto de vista de nuestro *autor*, la racionalidad humana es básicamente *instrumental* y los valores, aunque los agentes no sean concientes de ello, son sólo expresiones de su juicio positivo sobre aquellos bienes que les resultan útiles como *medios* para sus fines e intereses. Creo que para Gauthier ambas explicaciones son plausibles y cada una tiene su ámbito válido de aplicación. Sin embargo, sólo la segunda de ellas apela a la raíz última de la valoración, en apariencia no-instrumental, que hacen los agentes de su participación en actividades sociales y, por esto, sólo ella es la explicación verdadera o, mejor, la que posee el peso de una auténtica

explicación científico-causal. E, insisto, todo ello ocurre a espaldas y aún a despecho de lo que puedan pensar, percibir o afirmar expresa y sinceramente los agentes mismos, que son quienes hacen dicha valoración.

Pienso que nuestro autor aplica este mismo esquema explicativo —el cual, repito, sería consistente con el aspecto reduccionista de su propuesta—, para dar cuenta del hecho de que los agentes humanos valoramos muy especialmente y, al parecer, de manera intrínseca, aquella forma de moralidad a la que Gauthier se refiere como “afectiva”. Esto es, la moral propia de quienes deciden con base en sentimientos tales como la compasión o la vergüenza, o a partir de intereses “tuistas”, como es el caso de la preocupación por el bienestar de otros agentes. Si bien la moral por acuerdo es, según el autor, la única moral racionalmente fundada y, por tanto, sólo ella puede ofrecer una propuesta de moralidad capaz de convencer a un agente propiamente racional, sin embargo, también es innegable el hecho de que los agentes humanos, en general y en su día a día, suelen otorgarle un gran valor a los sentimientos como guías de sus decisiones morales. Del mismo modo, también parecen concederle un valor intrínseco al hecho de que haya personas que se muevan a actuar por esos sentimientos. Pero este fenómeno que, repito, la experiencia cotidiana parece mostrarnos de manera incontestable, se debe, no obstante, según Gauthier, a que de forma análoga a como ha ocurrido con el hecho de que valoramos *intrínsecamente* la *participación* o las actividades sociales/cooperativas, también los *sentimientos morales* han sido muy *útiles*, a lo largo de nuestra *evolución* como especie, para motivarnos y comprometernos a actuar éticamente (pp. 326-329).

De nuevo, creo que podemos, entonces, hacer la siguiente lectura de lo que nos dice aquí el autor: nos encontramos con dos explicaciones, una que proviene del punto de vista de los *actores morales*, y según la cual sentimientos tales como la compasión o la culpa son vinculantes y se deben valorar por sí mismos. Pero estaría la otra explicación, en mi opinión, más “profunda” a los ojos de Gauthier y que, por tanto, es aquella que proviene *del punto de vista del filósofo moral*, en este caso él mismo, que es quien conoce las auténticas causas a las que obedece este fenómeno de los sentimientos morales. Según esta segunda y verdadera explicación, la moral afectiva es un medio útil para la motivación y por eso ha sido tan exitosa desde una perspectiva evolucionaria. De allí que no debemos olvidar, insistiría Gauthier, que esta forma de moralidad *no* puede

proporcionar el fundamento racional de lo moral, fundamento que debería obedecer a la lógica de una racionalidad instrumental y que, por definición, remite al autointerés de los agentes racionales. En este sentido, para nuestro autor, los seres humanos, si bien *no* somos propiamente *market-selves*, no tenemos más remedio que admitir que el origen *último* de la valoración intrínseca que hacemos de aquello que Gauthier llama moral “afectiva”, está en nuestra racionalidad *maximizadora*.

Just as human beings may find participation with their fellows intrinsically valuable, even though the participatory activities have a non-tuistic rationale and explanation in their weakness and non-self-sufficiency, so human beings may find morality affectively engaging for its own sake, even though their moral principles have a non-tuistic rationale (pp. 326-327).

Tal vez en esta parte de la exposición ya el lector habrá advertido un interesante parecido entre la actitud que el filósofo canadiense asume ante la motivación moral propia del “hombre común” y la de algunos utilitaristas, especialmente la de ciertos autores contemporáneos.¹¹⁴ Como ya se ha mencionado, Gauthier es claro en separarse del utilitarismo, pero creo que en este punto coincide con dicha doctrina, al asumir una postura que algunos críticos, como A. Sen y B. Williams, califican de “elitista”. Esta postura se introduce subrepticamente, según dichos autores, en el caso del utilitarismo, en la tesis según la cual aquello que conviene que hagan o piensen que deberían hacer los agentes morales, no coincide con la *verdadera* explicación de eso que hacen o piensan que deben hacer dichos agentes; explicación que sólo conoce el filósofo moral “ilustrado”¹¹⁵, *i.e.* utilitarista. Sólo éste sabe la utilidad a la que obedecen las actitudes de la moral común o no ilustrada y por eso también sabe que, desde ese conocimiento suyo, desde su posición ilustrada y “desencantada”, puede darse cuenta del verdadero fundamento de la moral y, por tanto, puede ofrecerse una moral filosóficamente justificada. Y ello, aunque ésta sea desestimada por el hombre común e, incluso, le escandalice y, lo que es más importante: aunque lo socialmente útil sea,

¹¹⁴ Véase, por ejemplo, Smart (1988), y Harsanyi (1982).

¹¹⁵ Véase la crítica de Sen y Williams al utilitarismo (1982, pp. 1-21), y Williams (1972 y 1988). Esta tesis creo que se puede observar no sólo en el utilitarismo sino, en general, en filosofías que suscriben el proyecto reduccionista de explicar la moral como un epifenómeno de realidades no morales. Véase, además de autores destacados del utilitarismo contemporáneo, como Harsanyi (1982, 1983 y 1985) y Smart (1988), los aportes de algunos emotivistas del siglo XX, como Mackie (1990) y Stevenson (1984), así como la crítica que MacIntyre hace de dicha postura emotivista (1987 y 1991).

paradójicamente, que el hombre de la calle, lejos de convertirse en un filósofo utilitarista, siga guiándose por la moral tradicional. Creo que ya el lector habrá advertido que un crítico que haya seguido la discusión de Gauthier en contra de las sospechas escépticas que los sofistas manifestaban frente a la moralidad, bien podría convenir en que en este punto nuestro autor, a pesar suyo, se vería abocado a contradecir su propia réplica en contra de la postura de Trasímaco, pagando así el precio de asumir esta tesis elitista frente a la moral. Esto es, la tesis según la cual las convicciones morales que conviene que asuman las personas que no pertenecen al selecto grupo de filósofos, no deben confundirse con el auténtico conocimiento de las razones para dicha conveniencia, el cual, repito, sólo le pertenece al filósofo moral, en este caso, instanciado en el propio Gauthier.

2.5 La prelación del fundamento maximizador frente a los sentimientos morales ¿Cómo asegurar una auténtica motivación moral para el “nuevo” *homo oeconomicus*?

Tal vez previendo este tipo de críticas, Gauthier intenta mostrar una versión menos suspicaz de la relación entre lo que él llama la moral “afectiva” y la moral racional tal y como él la propone. Nuestro autor dice estar de acuerdo en que los agentes humanos tenemos una innegable capacidad para la moral afectiva y que esto ha tenido consecuencias manifiestamente positivas en cuanto a paliar las tendencias asociales de los individuos. Todo lo cual se debe, precisamente, al genuino carácter “tuista” de los intereses que motivan a quienes se preocupan sinceramente por el bienestar de sus semejantes y se dejan llevar por sus sentimientos de compasión o simpatía/empatía. Pero justamente esta es la motivación que queda por fuera de las posibilidades del *homo oeconomicus* y, aunque los agentes humanos reales no seamos yoes-del-mercado, sin embargo no hay que olvidar, advierte Gauthier, que si se quiere fundamentar racionalmente la moral, no puede apelarse al tuismo de la moral afectiva. En este punto creo que el lector se preguntará si el resultado de esta separación tajante entre los sentimientos morales, por una parte, y el fundamento racional de la moral, por la otra, ¿acaso no supondría una dicotomía insalvable, ante la cual todo agente humano no tendría más remedio que elegir radicalmente, sin fórmula de mediación, entre ser un sujeto pre-ilustrado y tuista que toma por guía sus sentimientos morales o, por el

contrario, ser un convencido de las tesis de Gauthier y, consecuentemente, tratar de aplicar el criterio maximizador y autointeresado en su vida diaria? Si así fuera se produciría o bien una suerte de esquizofrenia, o bien un desgarrador conflicto moral al interior de la conciencia de cada agente. Conflicto que le llevaría a tener que elegir, repito, sin matices ni fórmulas de mediación, entre actuar a partir de sus sentimientos morales y haciendo caso omiso de cualquier cálculo racional o, por el contrario, actuar reprimiendo esos mismos sentimientos y tratando de llevar a cabo los cálculos de un aplicado maximizador, caso en el cual actuaría únicamente a partir de lo que dichos cálculos le muestren como conveniente a sus intereses.

Suponiendo que tan desgarradora elección pudiera darse, a lo mejor se produciría, a su vez, una división entre dos tipos de agentes morales: aquellos que intenten decidir según la moral por acuerdo, y aquellos otros que encuentren muy extraña, artificiosa o hasta dolorosa o malsana dicha postura y, por tanto, prefieran permanecer fieles a su moral “tuista” de toda la vida. Esto es, se den cuenta que no se identifican y, a su entender, por buenas razones, ni con el *homo oeconomicus* ni con su acaso “extraño” tipo de motivación moral. Dado lo cual optarían por seguir comprometidos —y sinceramente— con una moral no-filosófica, tal y como lo han hecho hasta ahora. Por tanto, también harían caso omiso de una propuesta como la de Gauthier, por considerar que ésta iría dirigida a agentes como el *homo oeconomicus*, personaje que a los ojos de quienes decidan permanecer fieles a su moral “tuista”, posiblemente podría aparecer como una ficción, o tal vez como un ser demasiado excéntrico, e incluso como la víctima de algún delirio en virtud del cual ha perdido los vínculos con la realidad compartida por el resto de sus congéneres.¹¹⁶ Ante el peligro de este tipo de objeciones, creo que Gauthier ensaya de nuevo una contrarrespuesta ingeniosa e interesante: intenta mostrar que lo que explica nuestra capacidad afectiva para moral es que *previamente* hemos *reconocido las razones* para la moral. Así, según nuestro autor, la motivación afectiva para la moral ha requerido, como su condición de posibilidad, un previo convencimiento racional frente los principios morales.

A person who possesses a sense of duty displays an affective capacity for morality; he is moved to do his duty because it is his duty. A person who possesses a sense of justice

¹¹⁶ Véase, al respecto, la crítica de Strawson (1995) a este tipo de posturas en las que se asume que la actitud “objetivante” o “no participativa” sería la “más racional”.

also displays an affective capacity for morality; she is moved to do what is just because she considers it. Note that an affective capacity for morality presupposes a prior conception of morality; one cannot be moved by a sense of duty unless one antecedently believes some action to be one's duty [. . .] the just man is the person who, recognizing a certain course of action to be just, finds her feelings engaged by that recognition and so finds herself moved to adhere to that course of action because of its justice. To be sure, she may be moved in other ways as well, but for the just person, this motivation is essential (pp. 327-328).

Esta última afirmación, independientemente de lo discutible que pueda ser, al sostener como evidente que lo cognitivo y lo afectivo poseen naturalezas inconfundibles y separadas, así como una precedencia de lo primero sobre lo segundo, sin embargo creo que resulta muy interesante porque señala un terreno –hasta ahora no mencionado e incluso negado por el propio Gauthier–, terreno en el que el filósofo tiene que establecer qué tipo de lazos, si los hay, conectarían a las razones con los sentimientos involucrados en la moralidad. Si bien anteriormente nuestro autor había marcado una escisión insalvable entre ambas dimensiones, descalificando por completo la apelación a los sentimientos morales y asignándole el peso filosófico únicamente al fundamento racional –i.e. maximizador– de lo moral, ahora parece que Gauthier ve con mayor claridad que ello puede hacer peligrar la *genuina motivación moral* de su agente. Pues, si quiere que éste ya no sea pensado como el *homo oeconomicus* en su versión menos simpática, entonces tendría que asegurarse de que su agente modélico –ahora “mejorado” o “reestructurado”– tenga una motivación moral auténtica y estable. Esto es, que su cumplimiento de normas morales se dé por mor de la moral misma y no en aras de la mera utilidad que pueda obtener de ésta. Recordemos que el *homo oeconomicus*, tal como primero lo mostró Gauthier, si bien podía respetar las restricciones morales, lo hacía porque veía en ellas un medio útil para sus intereses. Pero precisamente por esto mismo presentaba el peligro de que podría hacer caso omiso de la moral en aquellas circunstancias en las que dicha trasgresión le prometiera mayores beneficios de los que obtendría si cumpliera con las normas de justicia.

Ahora bien, creo que en este “nuevo” terreno ahora descubierto por Gauthier, terreno en el que aparecen tanto las razones, como los sentimientos de los agentes morales, tal vez podría estar la salida a algunas de las perplejidades con las que nos hemos visto confrontados en este recorrido en el que se han ido examinando los argumentos del filósofo canadiense. Pues las dudas provocadas por la figura del *homo*

oeconomicus tal vez sólo pueden salvarse mostrando un agente que se comprometa con la moral en virtud de convicciones que le vinculen a ella tanto afectiva, como racionalmente. O mejor, un agente en el que lo afectivo y lo racional hagan parte de su –única– motivación para actuar moralmente: un ser que no tenga que pensarse como escindido o puesto ante la disyuntiva de elegir entre actuar llevado por convicciones o sentimientos ajenos a los cálculos de utilidad, o actuar únicamente a partir de dicho cálculo. A estas alturas de su discurso acaso Gauthier sea más conciente de que si desea hacer plausible su propuesta de una moral por acuerdo, ésta debe ganar la aceptación cualificada de *cualquier* agente racional, y no sólo de aquellos individuos que, en caso de que existan, se sientan plenamente identificados con el *homo oeconomicus*, figura a la que, a lo mejor y tal como lo hemos sugerido arriba, muchos agente sensatos puedan calificar de artificial. Además de ello, dicha aceptación, insisto, tendría necesariamente que incluir fuerza motivacional y no sólo razones de cálculo económico, para que permita modelar una figura plausible del agente moral-racional.

De allí que nuestro autor proponga la clave para ofrecer esta fuerza motivacional sin perder la imparcialidad y neutralidad propias de la moral por acuerdo: el agente al que iría dirigida su propuesta sería aquel que demande una racionalidad *no-tuista* para justificar las restricciones morales, pero que *también* tenga la capacidad para vincularse *afectivamente* con este tipo de restricciones (p. 329). Este tipo de agente no presentaría las desventajas de la parcialidad o falta de neutralidad de quienes, según Gauthier, sólo son sensibles a morales tuistas; pero tampoco tendría los inconvenientes del *homo oeconomicus*, a quien sólo mueve su propio interés y por ende, puede usar en su favor las restricciones morales o puede usar la moral como “fachada” (y uso acá los términos del mismo Gauthier; p. 329). Creo que con esta fórmula, a la que, por ahora, me permito calificar de “mixta”, nuestro autor parece querer unir las ventajas de una auténtica motivación moral, con las ventajas de una fundamentación racional de la moral, mediante una suerte de “mezcla” –aunque Gauthier no usa estos términos– entre una moralidad “à la Hume” y otra “à la utilitarista”. A los ojos de nuestro autor, y aunque parezca insólita, esta fórmula que, repito, uno podría considerar como “mixta”, sí resultaría viable dado que la auténtica motivación afectivamente vinculante posee una naturaleza a la que él se ha referido como “secundaria”. Es decir, como algo que “viene

luego” de que surge de manera “primaria” el verdadero fundamento de lo moral, fundamento cuya naturaleza es racional-maximizadora.

No obstante lo anterior, habría que admitir, reconoce Gauthier, que esta condición para la motivación moral no nos pone a salvo del peligro de que la moral pueda ser usada, de hecho o en escenarios reales, como instrumento de dominación (p. 329). En mi opinión, esto último nos lleva de nuevo a algunas de las perplejidades que se nos han presentado. Quienes utilicen de esta manera la moral estarían ejerciendo un modo de agencia que es, en principio, posible y del que sin duda podrían mostrarse ejemplos empíricos; modo de agencia que caracteriza al *homo oeconomicus* y que tantos problemas parece haberle planteado a nuestro autor, sobre todo si recordamos su alegato en contra de la tesis sofista. De allí que en este punto me permita insistir en que esta posibilidad de ejercer este modo de agencia y de usar la moral en la que otros creen de forma sincera, con el fin de manipularles ideológicamente, nos lleva de nuevo a ver la pertinencia de una de las críticas ya mencionadas en contra de las tesis de Gauthier; crítica que, a mi entender, seguiría en pie y pondría, por tanto, en entredicho, su proyecto de mostrar que los agentes humanos nos decidimos a ser morales por motivos autointeresados. Me refiero a la ya comentada objeción de que puede afirmarse, a la luz de los problemas que se le han señalado a la figura del *homo oeconomicus*, que aquello que garantizaría que cumplamos con la moral es que realmente *no* seamos *market-selves*. Por lo tanto, y es esto lo que me interesa concluir en este punto: pienso que resultaría bastante problemático establecer, como lo intenta hacer ahora Gauthier, una relación de *precedencia*, una relación del tipo fundamento-fundamentado, entre una supuesta “base” o “raíz” racional *i.e. instrumentalmente* racional, y una supuesta motivación “secundaria” de carácter afectivo, motivación que se daría “posteriormente” o “sobre la base de” dicho fundamento racional. Veamos, en lo que sigue, cómo nuestro autor intenta salir al paso de las dudas que pueda despertar esta objeción.

3. La insalvable condición de escasez y sus ineludibles consecuencias: maximización y moralidad

Ante las anteriores críticas y, sobre todo, ante las dificultades que sigue planteando el compromiso, irrenunciable para Gauthier, de fundar la moral en la racionalidad estrictamente *maximizadora*, nuestro autor intentará demostrar y, en mi

opinión, de una manera bastante ingeniosa, la tesis de que los seres humanos sencillamente no podemos escapar a nuestra condición de maximizadores y, por tanto, cualquier estrategia a la que se acuda para dar cuenta de la moral, debe tener su *último anclaje* en dicha condición. Para ello Gauthier (pp. 330 y ss.) se vale de una ficción de B. Suits (1978), como una metáfora que ilustraría el hecho de que la *escasez* es tan ineludible y definitiva para nuestro modo de ser humanos que, por *una parte*, determina que la forma de agencia que nos caracteriza como especie incluya necesariamente la conducta *maximizadora*. Por *otra parte*, la escasez hace tan imprescindibles las restricciones a dicho tipo de conducta, esto es, hace tan necesaria la *moral* que ésta, amén de no ser en absoluto un mal necesario para los hombres, es para ellos una fuente de bienes a los que sólo se podría acceder a través suyo.

En el cuento de Suits, tal como lo narra Gauthier, se recrea la historia de un mundo en el que no se presenta esta circunstancia de la escasez y de allí que cada habitante de ese mundo, llamado “Utopía”, pueda dedicarse a la tarea que le venga en gana e, incluso, pueda desarrollar muchas actividades distintas sin que con ello se cause algún problema a la producción de aquellos bienes que la gente necesite para satisfacer sus necesidades. Al no haber escasez las diversas actividades humanas no tienen un valor instrumental. No van dirigidas a la obtención de bienes escasos ni a la solución de problemas reales ni acuciantes.¹¹⁷ Es más, todo lo que hacen los habitantes de Utopía tiene la lógica de un juego; algo que se hace por diversión, por matar el tiempo, por dar rienda suelta a la expresión de los talentos pero que en sentido estricto no tiene otro fin distinto al de tomar parte en esas mismas actividades. Allí ocurre lo mismo que sucede con los juegos que jugamos acá en el mundo real; juegos cuyo único fin es el placer de jugarlos o cuyo sentido está simplemente en poder participar en ellos. Si en nuestros juegos se busca resolver ciertos “problemas”, de ese mismo modo se dan las cosas en Utopía, pero con la gran diferencia de que allí la lógica del juego se aplica a *todas* las actividades que desarrollan sus habitantes. Éstos se auto-imponen ciertos obstáculos que intentan solucionar mediante todos sus esfuerzos, quehaceres y prácticas, por lo cual, en su caso, las dificultades o los retos a superar son en realidad pseudo-problemas que ellos

¹¹⁷ El lector habrá advertido la similitud entre este cuento y la metáfora a la que acude Hume para mostrarnos cómo sería un mundo en el que no hubiera escasez y las personas no necesitaran esforzarse por conseguir bienes escasos. Cf. Hume, *Investigación sobre los principios de la moral*, sección 3, Parte I, §§ 13-19.

mismos ha creado por el mero placer de intentar sortearlos. Empero, ocurre que un buen día los habitantes de Utopía caen en cuenta de que convendría asignarle un valor *instrumental* a esos juegos que allí se juegan pues, de lo contrario, la gente va a terminar por desanimarse al ver que invierte toda su vida en meros juegos que no van dirigidos a obtener, finalmente, nada, aparte del placer de jugar. Así, con el tiempo se llega a olvidar en Utopía que las actividades humanas fueron alguna vez consideradas como meros juegos y se termina por concederles un valor *instrumental*, el cual es inseparable del valor *intrínseco* que acaba asignándosele a los *bienes* obtenidos mediante dichas *actividades*.

La ficción de Utopía le sirve a Gauthier de contraste, al modo del negativo de una fotografía, para mostrar que nuestra situación es exactamente la opuesta. Pues en nuestro caso, según el canadiense, no podemos darnos el lujo de ignorar la escasez y, por tanto, nuestras actividades acá, en el mundo real, necesariamente deben dirigirse a obtener *algo* con ellas. Esto es, nosotros, a diferencia de los ciudadanos de Utopía, no podemos asumirlas como juegos, ya que mediante ellas buscamos solucionar problemas reales y acuciantes: problemas auténticos, nacidos de nuestras necesidades, que también son reales y acuciantes. Si en Utopía el valor de las actividades humanas reside solamente en el hecho mismo de poder participar en ellas, en cambio, en el mundo humano real esa participación se hace por mor de aquellos bienes que se van a obtener mediante dichas actividades. Para nosotros, a diferencia de lo que sucedería con los habitantes de Utopía, las tareas que desarrollamos no son meros juegos y los problemas que se intenta resolver mediante ellas no son pseudo-obstáculos autoimpuestos con fines lúdicos. Por el contrario, repito, son auténticas dificultades nacidas de unas necesidades nuestras que nos apremian, lo cual se complica aun más si tenemos en cuenta que para satisfacer dichas necesidades contamos con recursos escasos; recursos que *no* son para nada prescindibles y que sólo podemos obtener mediante lucha y esfuerzo.¹¹⁸

Por lo anterior, según Gauthier, en nuestro mundo humano, que es la contracara de Utopía, pasa exactamente lo opuesto a lo que ocurre allí. Acá hemos llegado a apreciar tanto esas actividades dirigidas a solucionarnos problemas reales y

¹¹⁸ Gauthier (pp. 332) afirma que si en *nuestro* mundo, al igual que ocurre en Utopía, también hay juegos, por contra de lo que sucede allá, aquí les asignamos un valor *intrínseco*, *no instrumental*, lo cual se explica precisamente porque ellos son para nosotros unos islotes de descanso en un mar de actividades que nos cuestan trabajo y que se realizan por su valor *instrumental*, esto es, como medios para obtener los bienes que necesitamos para nuestra vida y no simplemente para divertirnos o matar el tiempo.

proporcionarnos bienes escasos, que hemos terminado por asignarles un valor *intrínseco*, llegando incluso a considerar como valiosos por sí mismos a la lucha y al esfuerzo (*seeking and striving*) invertidos en la obtención de tales bienes; con lo cual a la postre hemos olvidado su original valor *instrumental*. En todo caso, la idea de una sociedad humana como la de Utopía, esto es, signada por la plenitud y no, como la nuestra, determinada por la escasez es, según Gauthier, una quimera que, a su vez, está asociada a la ficción de que podemos liberarnos de aquellas actividades mediante las cuales se obtienen esos bienes que tanto necesitamos y que sólo podemos lograr gracias a nuestros esfuerzos y luchando contra dificultades reales.

3.1 La imposibilidad de la utopía marxista. Nuestra esencia maximizadora y la necesidad de la moral no sólo como restricción, sino también como medio

Como ya lo habrá previsto el lector, Gauthier (pp. 333-334) le atribuye esta idea utópica a Marx, quien vaticina el fin de la era capitalista y la llegada del comunismo y, con él, el final de la división del trabajo tal como la conocemos, esto es, la liberación de las ataduras que tienen los trabajadores con un único oficio que termina por empobrecerles y atrofiarles. Según Marx, citado por Gauthier, en la sociedad capitalista cada individuo se dedica por completo a un solo oficio, cerrándose para ella o él la posibilidad de desempeñar diversas tareas y desarrollar sus talentos y capacidades e, incluso, no pudiendo siquiera elegir entre distintos trabajos según sus gustos y aptitudes. Pero con la llegada del comunismo todos tendremos la oportunidad de elegir las tareas que queramos, pudiendo así desarrollar diversos talentos. El cumplimiento de esta utopía marxista requeriría, según Gauthier, que no hubiera escasez, ya que ella es la condición que determina la necesidad de que cada cual se dedique a una y sólo una labor productiva. Pero el caso es que, para nuestro autor, es la escasez, y no plenitud anhelada por Marx, la circunstancia que determina indefectiblemente nuestra humana condición. Por este motivo no está a nuestro alcance el poder escapar a dicha necesidad, lo cual implica que no podemos liberarnos de la división del trabajo tal y como la conocemos ni, por tanto, tampoco de las ataduras que cada quien tiene con el oficio que desempeña, ni con el consiguiente cierre de opciones para que se dedique a otras tareas o desarrolle todos los talentos que desee.

Lo anterior no sólo afecta, según el canadiense, la producción de bienes *materiales* puesto que, aún si ocurriera que éstos no fueran escasos, de todos modos sí lo serían las *capacidades*, las *energías* y el *tiempo* que tienen las personas. Esta segunda forma de escasez, la escasez “de consumo” (*scarcity of consumption*), impediría que cada individuo pudiera humanamente desarrollar todos los talentos o tener todas las fuerzas y todo el tiempo que quisiera para desempeñar diversos oficios. Por lo tanto, la escasez, según Gauthier, hace necesario que cada cual se dedique por completo y durante toda una vida a una tarea y escoja una sola forma de vida, renunciando a otras, no sólo por la necesidad de producir bienes materiales escasos, sino también por la necesidad de que las tareas se puedan desempeñar adecuadamente. No sería factible que por la mañana seamos cirujanos y por la tarde periodistas; un día seamos albañiles y al día siguiente abogados. Ninguna persona, por genial y eficiente que sea, puede realizar en sí misma todas las formas de vida ni llevarlas adelante con la maestría necesaria. Es, entonces, la *escasez y no la moral*, según nuestro autor, el auténtico *mal necesario* que nos obliga a elegir hacer una cosa y dejar de hacer otras (p. 335).¹¹⁹ Y si la *necesidad de elegir* es determinante para nuestro humano modo de agencia, esto trae la importante consecuencia de que *estamos avocados necesariamente a ser maximizadores*: seres que no tienen más remedio que estudiar sus opciones y elegir aquella que les resulte la más “económica”; la más eficiente y eficaz para satisfacer sus necesidades y deseos; la vía menos costosa y más factible, dadas sus preferencias, sus posibilidades alcanzables y las restricciones empíricas que determinan su contexto de acción.

¹¹⁹ Creo que conviene advertir que un lector que quisiera revivir las preocupaciones de Marx —y no sólo de éste—, tal vez podría señalarle a Gauthier algo que nuestro autor parece pasar por alto y que posiblemente siga siendo motivo de válida preocupación para cualquiera que se pregunte por las condiciones para la libertad humana. A saber, la *imposibilidad* que tienen muchos de *elegir* en escenarios sociales reales, incluyendo de manera evidente a nuestra moderna sociedad de mercado. En estos escenarios la *maximización* que impone el sistema de *mercado* no necesariamente va acompañada de la posibilidad de *elegir* oficios o formas de vida sino que, más bien, suele asociarse al hecho de que todavía muchos agentes *no tengan opciones*, no tengan acceso a oportunidades de elección que les permitan desarrollarse y enriquecerse en sus capacidades y talentos humanos. Ello, no sólo por condiciones de privación material, sino también por falta de acceso a opciones que están desde un principio vedadas para muchos en razón de su origen socio-económico y de diversas formas de discriminación y desventaja que, o bien acompañan la falta de riqueza material, o bien van más allá de ella, v.g. los problemas de género, los problemas raciales. Ante las múltiples formas de exclusión que hacen imposible para muchos tener esa supuesta posibilidad de elegir, creo que resultaría —y, repito, no sólo a los ojos de un marxista—, entre ingenuo y cínico el afirmar, como lo hace Gauthier, que “estamos avocados a elegir” —todos, y no sólo algunos privilegiados— entre diversos oficios y formas de vida. Nuestro autor, sin embargo, no parece caer en cuenta de esta crítica y concluye que en materia de talentos y realización personal también estamos “destinados a escoger”, dado el ineludible problema de la escasez. (Cf. Sen, 1999, pp. 87-110; Young, I. M.; 1990, cap. II).

Por consiguiente, Gauthier considera un error el rechazo de la idea de que el ser humano es un maximizador. Para el filósofo canadiense, pensarnos a nosotros mismos como esencialmente maximizadores no sería en absoluto una forma limitada o empobrecida de concebirnos, como tampoco sería –en un sentido etnocéntrico– una mera expresión de nuestros particulares prejuicios como occidentales modernos. Por el contrario, la forma distorsionada o ilusoria de autorrepresentarnos sería, para Gauthier, aquella que se expresa en la ficción de Utopía. No obstante, el autor advierte que él tampoco cree que la necesidad de la maximización implique que el *homo oeconomicus* sea el ideal –normativo– de agencia humana, pues el “no-tuismo” que caracteriza a este tipo de agente sería tan inaplicable a nuestras condiciones, como lo es la idea de una plenitud de recursos materiales y humanos que hiciera innecesaria la maximización. Lo cual, a su vez, trae otra consecuencia de la mayor importancia: si necesariamente somos maximizadores, entonces no tenemos más remedio que sortear las dificultades que acarrearía la convivencia y la posibilidad de conflicto entre unos seres que buscan, cada uno, su propio beneficio. Es decir, se haría imprescindible acudir a mecanismos de arbitraje que permitan que estos seres, amén de no destruirse mutuamente, puedan maximizar su utilidad individual, pero de tal modo que los unos no obstaculicen la maximización que intentan llevar a cabo los otros.

Adicionalmente, es indispensable que cada uno pueda contar con la cooperación de sus congéneres, pues sólo de este modo podrían todos acceder a aquellos bienes que necesitan y que cada individuo no puede obtener únicamente mediante sus propios medios. Habría entonces que concluir que las ineludibles e inseparables condiciones de la *escasez* de recursos, la necesidad de *maximizar* y la *finitud* humana –que nos hace seres menesterosos, y también nos obliga a elegir una tarea y renunciar a otras– nos llevan a una doble necesidad: la de contar con unas restricciones *morales* a la conducta maximizadora, y la de *cooperar* con el fin de obtener aquel tipo de bienes que serían inaccesibles para nosotros si no contáramos con el esfuerzo conjunto de varios o muchos individuos. Para Gauthier, toda forma de vida humana se caracteriza, precisamente, por requerir de este tipo de bienes cooperativos y de allí que la finitud de cada agente, su menesterosidad, amén de no ser equiparable a ese mal necesario del que hablaban los sofistas sea, por el contrario, una prueba de que nuestra naturaleza es

esencial y positivamente sociable.¹²⁰ A su vez, la sociedad como empresa de cooperación, amén de sustentarnos, nos enriquece como individuos: cada cual contribuye con dicha empresa, al tiempo que se beneficia y “florece” gracias a los aportes de sus conciudadanos (si bien Gauthier, por supuesto, no usa el término “florecer”).

Taking scarcity for granted, as giving instrumental point to our activities, have we reason to regret that some music requires an orchestra, or that some games (since we have no reason to deny the value of games in a world that is not all game-playing) are for teams? [. . .] May we not rather suppose that our social capacity to find value in participating in one of the main sources of enrichment in human life? (p. 337).

3.2 El irrenunciable individualismo, el valor instrumental de la cooperación y, de nuevo, la base racional de los sentimientos morales

Gauthier usa su anterior argumentación como base para sustentar y, creo que nuevamente de una manera bastante ingeniosa, su tesis (arriba comentada) de que los seres humanos, por razones ancladas en nuestra evolución como especie, hemos llegado a otorgarle un valor tan prominente a las acciones y prácticas *participativas*, que con ellas ha terminado por ocurrir lo mismo que en el caso de las actividades *productivas*. Es decir, hemos olvidado su original valor *instrumental*, hasta llegar a atribuirles un valor *intrínseco*. Por esta misma vía, según nuestro autor, puede explicarse la especial valoración que hacemos de nuestros congéneres: las actividades en las que nos involucramos con ellos y los beneficios que obtenemos de tales actividades participativas, suponen necesariamente el establecimiento de lazos entre las personas, lazos en virtud de los cuales cada uno de nosotros llega a apreciar enormemente a aquellos compañeros con los cuales se involucra y coopera. En razón de la ineludible circunstancia humana de la escasez de consumo, esto es, de la imposibilidad de que cada cual pueda desarrollar todos los talentos, realizar todas las tareas o producir todos

¹²⁰ En este sentido, Gauthier (pp. 335-337) dice estar de acuerdo con la valoración positiva que hace Rawls de la imposibilidad humana de que un agente solitario, por sus solas fuerzas, pueda desarrollar todos los talentos o llevar a cabo todas las tareas necesarias para obtener bienes que únicamente pueden lograrse mediante la labor cooperativa que lleven a cabo varios o muchos agentes de manera conjunta. Por ello, el canadiense acude al mismo ejemplo de Rawls: el sonido de una orquesta sólo puede obtenerse gracias a la unión de los talentos y esfuerzos de varios músicos, cada uno aportando la ejecución especializada o experta de su instrumento. Cf. Rawls (1999, § 79, nota 4).

los bienes que desee, hemos terminado por otorgarles un gran valor a aquellos productos, tanto materiales como culturales, que se obtienen de la cooperación; así como también a las virtudes, talentos y buena disposición que hallamos en las personas que de este modo cooperan y nos benefician. Todo ello nos revela, según Gauthier, lo lejos que tendríamos que estar de ver como males necesarios a la menesterosidad humana y a la necesidad de cooperar. Por el contrario, estas notas fundamentales de nuestra naturaleza y de las circunstancias en las que ésta se desarrolla, serían la fuente de aquellos bienes que más apreciamos para una vida propiamente humana, tanto material como culturalmente hablando.

Empero, el autor insiste en que la justa aprobación que se ha hecho de nuestra tendencia a valorar de manera especial las formas cooperativas de praxis, así como la compañía y el apoyo de nuestros semejantes, sin embargo no debe hacernos perder de vista que las actividades participativas originalmente tienen un valor *instrumental*. Del mismo modo, tampoco hemos olvidar la divisa individualista: es importante tener siempre presente que la razón de ser de dichas actividades es proveer bienes a los *individuos*. Para Gauthier, es el individuo lo que importa; es el bienestar y el desarrollo de éste lo que tiene que darnos la clave de aquello que habría que valorar en la vida pública y en las actividades de participación social. No hay que caer, entonces, aconseja nuestro autor, en la tentación de pensar que lo público y las actividades que lo caracterizan tengan un valor en sí mismos.

In valuing participation we appreciate the importance of public life. Our conception of society embraces the forum as well as the market-place. But just as value of seeking and striving requires these activities to have goals beyond themselves, so the value of participating requires that public life takes its rationale from the varied realizations of individuality that it makes possible. Public life considered apart from its instrumental roles would become yet another game (p. 337).

En vista de esta prioridad que ha de dársele al individuo y a sus necesidades, por encima de la vida pública, la cual según Gauthier tendría, repito, un valor instrumental relativo a los bienes necesarios para satisfacer esas necesidades individuales, el filósofo canadiense cree que debe hacerse, entonces, una valoración positiva de la cooperación humana y del hecho de que seamos seres sociales. Dicha valoración arroja el importante resultado, arriba ya mencionado, de que la moral, mediante la que se establecen unas reglas de juego que permitan que la cooperación vaya en favor de todos y de cada uno,

sin favorecer ni perjudicar a nadie en particular, debe ser vista como algo indispensable y deseable para que podamos desarrollarnos individualmente bajo el amparo de una sociedad. La moral, como ya se ha dicho, no sólo sería fuente de *restricciones*, lo cual es fundamental para mantenernos a salvo de las agresiones de otros agentes,¹²¹ sino también fuente de *posibilidades* para que cada uno pueda lograr sus fines y desarrollar las actividades en las que requiera involucrarse, contando para ello con una seguridad, una estabilidad y una riqueza incomparablemente mayores que aquellas que se podrían hallar en un supuesto estado de naturaleza. Pues en éste, según Gauthier, tal y como puede deducirse de los relatos de Hobbes y de los sofistas, los hombres vivirían bajo la dictadura de la fuerza, la coerción y la falta de posibilidades de desarrollo de sus talentos. Pero dado que siempre es posible pensar en un mundo así de riesgoso, nuestro autor advierte que la visión positiva que él nos presenta de la moral y de la vida en sociedad no es algo que resulte evidente de suyo para cualquier agente racional ni, mucho menos, que sea una idea aplicada de hecho en todos los escenarios sociales, o por todos los seres humanos. Esta idea sólo sería compartida por quien se considerara a sí mismo un cooperador MR y quisiera verse rodeado de otros cooperadores como ella/él, pero no lo sería por un agente al que le parezca más atractiva la opción de ser un maximizador MI, o se vea inmerso en un medio social que le muestre dicha opción como la más razonable.

Morality takes a different colouration when viewed in relation to participation. For asocial seekers and strivers morality could be no more than a needed but unwelcome constraint. But for those who value participation, a morality of agreement although still a source of constraint, makes their shared activity mutually welcome and stable, ensuring the absence of coercion or deception (p. 337).

Afirma entonces Gauthier, que aunque cualquier cooperador puede hallar buenas *razones* para *cooperar*, también es cierto que debe evaluar el *precio* de esta cooperación, al modo en que lo hace un maximizador al calcular los costes que tendría que invertir en el logro de un determinado fin. Así como el segundo ha de tratar de establecer si los *costes* invertidos son compensados por los *beneficios* obtenidos, de modo similar el primero debería evaluar si los sacrificios que tendría que hacer para poder disfrutar de la cooperación social, son realmente compensados por ésta. Si el

¹²¹ De nuevo, puede apreciarse la influencia de Hume, concretamente, en cuanto a su explicación de la necesidad de la justicia. Cf. Hume (Trad.1993, Sección 3).

cooperador advierte que sus potenciales copartícipes van a invertir unos costes menores en comparación con los que a ella/él le corresponderían, el resultado lógico es que le resulte muy poco razonable o atractivo cooperar con esas personas que intentan explotarle. Esto nos lleva de vuelta tanto al principio del MRC (*minimax relative concession*), como a la prioridad de la base “no-tuista” de la moral, prioridad que, según Gauthier, explica la característica del no-concernimiento mutuo, propia de los agentes morales.

En este punto, nuestro autor hace un llamado de atención para que esta idea de la necesidad de la cooperación, idea en la que él hace un énfasis especial al final de su texto, no vaya a hacernos perder de vista aquellos otros elementos fundamentales de la agencia humana en los cuales anteriormente Gauthier insistió tanto. De allí que, de nuevo, el filósofo canadiense resalte la relación de *precedencia* entre la base *racional* de la moral y las *afecciones “tuistas”* asociadas a esta última. Según Gauthier, antes de que desarrollemos valores “tuistas”, hemos debido *primero* verlos como *razonables*, *i.e.* favorables a *nuestros intereses*. De manera que la preocupación que mostremos por otros agentes y su bienestar, es *secundaria* o derivada de esas base *racional no-tuista*, en virtud de la cual se nos debe considerar seres fundamentalmente autointeresados y no-concernidos por otros agentes; seres que pueden desarrollar sentimientos morales y empatía ante sus congéneres, pero únicamente como algo derivado o fundamentado sobre una previa base racional –*i.e.* maximizadora y autointeresada–. Sólo si originalmente los agentes han reconocido dicho fundamento, pueden luego desarrollar sentimientos empáticos que primero les lleven a interesarse en sus congéneres y, finalmente, en la moral misma (p. 338).

If persons develop tuistic values primarily in the context of shared activities, then the acceptance of moral constraints underlies these values. In the absence of a morally acceptable division of benefits, persons will come to view their shared as exploitative rather than cooperative, and their fellows as having interests opposed to their own. And this must be inimical to the development of continuation of tuistic bonds. It might seem that, in admitting that persons do come to take an interest in the interests of their co-participants, we undercut the rationale for requiring that moral constraints have a non-tuistic basis. But if moral constraints underlie tuistic values, then they must have a basis independent of those values. Persons come to take an interest in their fellows because they recognize their mutual willingness not to take advantage of each other and to share jointly produced benefits on a fair basis. In accepting moral constraints they do not express their concern for each other, but rather they bring about the conditions that foster such concern. [. . .] Having first engaged their reason morality now engages their affections. An affective capacity for morality does not give rise to moral constraints but

presupposes their prior recognition; the desire to do one's duty cannot determine the content of duty. [...] Persons rationally recognize the constraints of morality as conditions of mutually beneficial co-operation. They then come to value participation in co-operative and shared activities that meet these constraints, and to take an interest in their fellow participants. And finally, they come to value the morality that first appeared to them only as a rational constraint (p. 338).

Creo que a un lector que haya seguido con atención los matices introducidos por nuestro autor en su último capítulo, esta nueva insistencia en que la racionalidad, *i.e.* el autointerés y el no-concernimiento por otros agentes, sería prioritaria frente a los sentimientos morales y frente al interés en el bienestar de las demás personas, podría parecerle un inexplicable paso atrás que nos llevaría de nuevo al *homo oeconomicus*, de cuya versión más dura Gauthier había intentado separarse. Por otro lado, también parecería que aquí el autor está proponiendo una génesis de la conciencia moral, al afirmar que primero se da un reconocimiento de la racionalidad de las restricciones morales, esto es, de su razón de ser para el propio beneficio de cada agente, y sólo posteriormente se produciría el desarrollo de los sentimientos asociados a la moralidad. Primero se tendría la necesidad racionalmente dictada –*i.e.* autointeresada– de cooperar con otros agentes y luego, y únicamente de una manera derivada, sobrevendría el aprecio emotivo por esas otras personas y por las actividades en las que se participa con ellas.

No obstante, Gauthier advierte que su propósito no es mostrar la *génesis* de la conciencia moral sino sólo hacer una *reconstrucción racional, no empírica*, de ésta.¹²² El autor insiste en que su idea no es entrar en polémicas sobre una *cuestión de hecho*, sino *reconstruir lógicamente*, en el sentido de *fundamentación racional*, el compromiso que podamos desarrollar con la moral y, aunque a algunos sorprenda, Gauthier afirma que su idea de una *prelación de lo racional sobre lo afectivo*, en vez de demeritar a esto último, más bien muestra su razón de ser. Pues, tal y como él ve el asunto, la reconstrucción racional de la moral es condición indispensable para que pueda darse razón de los *sentimientos* asociados a ésta, es decir, para que pueda apreciárselos como *justificados racionalmente*. Así, según nuestro autor, gracias a dicha reconstrucción

¹²² El filósofo canadiense (pp. 338-339) incluso reconoce que usualmente nuestro desarrollo como agentes morales suele ser narrado por los psicólogos en el sentido inverso al que él lo presenta, esto es, normalmente se afirma que primero, de niños, fuimos educados sentimentalmente en la moralidad y luego, al madurar, pudimos hacernos cargo de la racionalidad de los principios morales.

puede explicarse de qué modo estos sentimientos *no* deben considerarse como una suerte de disfunción de lo racional, entre otras cosas, porque, como ya lo ha reiterado Gauthier, el agente humano no es del todo un *homo oeconomicus*.

De esta manera, el filósofo canadiense cree estar haciéndole un favor, antes que perjudicarla, a la moral que él llama “afectiva”; ya que muestra la base racional que la justifica, que la precede y que explica por qué, en nuestro caso y a diferencia de lo que le ocurriría al hombre “natural” o al *homo oeconomicus*, la moral no es ese mal necesario al que se referían los sofistas (p. 339). Por lo anterior, Gauthier sostiene que, frente a otros intentos de fundamentación de la moral, su propuesta contractualista posee la ventaja comparativa de que no sólo puede demostrar la base racional de lo moral, sino que, además, logra hacernos ver cómo, a partir de esa base, y de una manera derivada, surgen los sentimientos morales. Si por esta vía puede darse cuenta de la razón de ser de tales sentimientos, entonces se obtiene el importante logro de que se demuestra que la existencia de éstos, amén de no contradecir la plausibilidad de la tesis de Gauthier, adicionalmente la refuerza, puesto que sólo desde ella puede verse con claridad de qué modo la moral por acuerdo, amén de fundamentarse sobre una base racional-maximizadora, incluye y explica desde sí misma ese otro elemento que algunos –erróneamente, en opinión de nuestro autor– podrían considerar ajeno o dicha base; a saber, esos sentimientos morales que ahora, mediante la demostración que hace aquí el filósofo canadiense, aparecen, según él, como racionalmente justificados (p. 339).¹²³

¹²³ Al igual que Gauthier, un utilitarista contemporáneo, como J. Harsanyi (1982, 1983 y 1985) quien, como arriba se comentó en un pie de página, también participa del proyecto reduccionista con el que simpatiza el autor canadiense, utiliza esta misma maniobra orientada a demostrar que la racionalidad maximizadora “fagocita” a los sentimientos morales y que tras ello hay una explicación evolucionaria. Si bien Gauthier, como ya se ha mencionado, asume una posición fuertemente crítica no sólo frente a Harsanyi, sino en general frente al utilitarismo y al bienestarismo económico, no obstante, a mi parecer, en este punto la argumentación de *La moral por acuerdo* es bastante parecida a la de Harsanyi, pues con ambas se busca hacerle ver al lector de qué forma los mencionados sentimientos, a los cuales uno tal vez podría considerar como el *quid facti* que desautorizaría este tipo de proyectos reduccionistas, sin embargo, y aunque pudiera parecer contradictorio, podrían y deberían explicarse como funcionalmente determinados por la racionalidad maximizadora. Y ello, sin desmedro de que, tal y como ambos autores lo reconocen, esta suerte de “alquimia” pueda aparecer como insólita a los ojos de un agente “pre-ilustrado” o “tuista” que considere a los sentimientos morales como un elemento ajeno e irreducible a la mera búsqueda racional, *i.e.* autointeressada y no solidaria, de la utilidad individual en el caso de Gauthier, o eventualmente solidaria en el de Harsanyi.

3.3 La crítica a los yoes discontinuos de Parfit. La capacidad crítica del agente liberal y su preocupación por sus yoes futuros

Luego de su anterior exposición de las diferencias entre las versiones clásicas del *homo oeconomicus* y su propia –y “reformada”– versión de dicha figura, Gauthier intenta mostrar cómo, una vez que a ésta se le han hecho los ajustes necesarios, es decir, una vez se ha producido en ella lo que antes me he permitido llamar una “transición” o “metamorfosis”, entonces debe arribarse al modelo del agente moral-racional con el que el autor afirma que se comprometerá de forma definitiva: su “individuo liberal”. Para establecer los rasgos fundamentales de este personaje, el filósofo canadiense comienza señalando lo que considera son algunos de los fallos presentados por otras versiones (distintas a la suya) de la agencia humana. Un buen ejemplo de estos fallos puede apreciarse, según Gauthier, en la tesis sostenida por D. Parfit (1984) de que los distintos *yoes* que cada uno de nosotros ha sido, en realidad son yoes *discontinuos*, a pesar de su aparente ligazón. De manera que, para Parfit, ese “yo” que cada quien se atribuye a sí mismo, no es otra cosa que el resultado de “agregar” dichos yoes, agregación que se lleva a cabo por razones de economía del lenguaje y otro tipo de necesidades pragmáticas, pero que no autoriza en absoluto el que se sostenga la idea tradicional de una unidad –propiamente dicha– que subsuma a esos distintos yoes. En contra de la tesis de Parfit, Gauthier cree que habría que asumir un compromiso “más fuerte” con una versión continua y más sustantiva del yo y de su *identidad* personal, versión que viene avalada gracias al hecho, innegable según el canadiense, de que el *yo presente* es capaz de pensar en sus *yoes futuros* y puede preocuparse –ahora– por las preferencias que esos yoes tendrán posteriormente.

Personal identity has an essentially practical dimension. The self at time t^1 is identical with the later self at time t^2 , to the extent it identifies with the later self, and this identification is measured by the weight given to the expected preferences of the self at t^2 in the preferences at t^1 (pp. 342-343).

Nuestro autor cree que a su agente liberal se lo debe pensar bajo el modelo de un yo que, a diferencia del que propone Parfit, sí incluye de manera fundamental la capacidad que un agente humano tiene para mantener una *identidad* personal a lo largo del tiempo e *integrar* sus *yoes sucesivos* en un *continuo*. Según Gauthier, esto queda

probado, repito, gracias al hecho de que todo yo actual puede tener en cuenta, para sus preferencias de *ahora*, aquellas otras preferencias que tendrá en el *futuro*, lo cual indica que el agente liberal es reflexivo hasta tal punto que puede *examinar* de forma *crítica* sus preferencias y elegir por sí mismo aquello que considera bueno para ella/él. En otras palabras, se trata de un agente *autónomo*, a quien se le pueden atribuir preferencias *consideradas* (pp. 345-346). Éstas constituirían, según Gauthier, la prueba de que estamos hablando de un agente *racional* no solamente en el sentido de llevar a cabo un *cálculo maximizador* orientado a satisfacer sus actuales preferencias, sino también en el sentido de ser *completamente (fully)* racional, es decir, *reflexivo* y autocrítico en cuanto a sus preferencias futuras. Dicha reflexión se traduce en el hecho de que el agente puede ver sus futuros yoes *en conexión* con su yo-presente y advertir que esta conexión entre el yo del momento t^1 y los yoes de los momentos t^2 , t^3 y t^n no sería ni arbitraria ni enteramente accidental, lo cual dispone al yo-actual a *identificarse* con sus futuros yoes y a concederles, en sus actuales preferencias, un peso importante a las preferencias que tendrán esos futuros yoes. El agente asume que estas últimas se derivarán de sus preferencias actuales gracias a que, precisamente, puede adelantar ese proceso de reflexión, el cual involucra también el aprendizaje sobre las experiencias que vaya teniendo a lo largo del tiempo.

Sin embargo, esta previsión que el yo-actual hace de sus futuros yoes está marcada por la incertidumbre, dada la ignorancia que tiene el primero sobre esas preferencias que supone tendrán los segundos. De allí que cada agente trate de domar esta incertidumbre al momento de elegir las normas e instituciones que afectarán a sus yoes futuros, lo cual muestra, según Gauthier, la importancia que tiene la elección arquimédica de tales normas e instituciones. Pues éstas deberán ser capaces de responder a la diversidad de preferencias que puedan presentarse por parte de los futuros yoes de cada agente y de allí el enorme atractivo que tiene, para nuestro autor, una postura liberal que, como la suya, permite que se abra ante los individuos un amplio abanico de opciones para su elección de diversas formas de vida. En el corazón de esta postura liberal está la importancia fundamental que se le concede a la autonomía de los agentes: para un liberalismo como el que él propone, según Gauthier, es clave partir de la base de que cada individuo está capacitado para reflexionar y autocriticarse sobre sus razones para actuar. Esto significa que al agente liberal se lo debe pensar como alguien

que *no* está atado sin remedio a sus actuales preferencias ni, mucho menos, a las preferencias de alguna instancia externa a ella/él, v.g. otros agentes o un colectivo social, si bien es innegable que, para poder elegir, el individuo tiene que partir de unas preferencias que son un *factum* y un necesario punto de partida. Este último, sin embargo, siempre podrá ser objeto de revisión y de crítica, de modo que sea susceptible de sufrir cambios, incluso radicales, a lo largo del tiempo.

De esta interesante y, creo yo, muy atractiva defensa que hace Gauthier de una conexión, en palabras suyas, “fuerte”, entre los distintos yoes que va teniendo un agente a lo largo del tiempo, quisiera resaltar en *primer* lugar, el hecho de que dicha reflexión sea *únicamente* sobre yoes *futuros* y no involucre *también* un examen crítico de los yoes *pasados*. En *segundo* término, a mi entender, no se ve con claridad *qué es lo que permite* que un yo-presente se identifique con un yo-futuro, o *en virtud de qué* ocurre tal identificación, lo cual nos remite, *finalmente*, al problema de que, en mi opinión, tampoco se muestran cuáles son los *referentes normativos* a los que acude el yo para evaluar o criticar sus preferencias, tanto las del futuro mencionado por Gauthier, como las pasadas, a cuya evaluación nuestro autor no parece otorgar un peso en la reflexión del yo actual. Pienso que resulta bastante acertado, por parte de Gauthier, el que señale la importancia que tendría, para una noción más fuerte y sustantiva del yo, la preocupación del sujeto por sus *yoes futuros*. Así, creo que estos últimos admitirían que se los viera como algo *más* que meros *hechos* futuros sobre los que el agente –meramente maximizador– simplemente se limitara a imaginar, prever y planificar estratégicamente (sin que con ello se pretenda en absoluto negar la importancia de dichas imaginaciones y planes), siendo, por tanto, la reflexión sobre dichos yoes futuros una parte fundamental de aquello con lo cual el agente está interesado en comprometerse y en comprometer su vida, y sus planes a futuro. Lo cual también implicaría un sentido para la ya estudiada autorrestricción a la que se somete el agente, mediante su compromiso con las normas que le permitan una convivencia más armónica con otros agentes.

La apelación a estos yoes futuros, pues, creo que resulta importante para dotar al agente de Gauthier de una sustantividad y de un carácter reflexivo que se echaba en falta en gran parte de su texto. No obstante, y aún a pesar, repito, de la acertada referencia que hace nuestro autor al peso que un agente liberal tendría que concederle a

tales yoes futuros, creo que el significado y las implicaciones de dicho “peso” resultan no poder ser desarrollados del todo dentro de su argumentación, dado un lastre que acarrea su agente modélico: su carencia de un contexto y de unos referentes normativos, tal y como se vio en el primer capítulo. Así, en la descripción que hace Gauthier no se ve cómo se desarrollaría la mencionada evaluación de los yoes futuros, puesto que no resulta fácil establecer a partir de qué criterios o referentes normativos es que ésta podría llevarse a cabo. Esto nos lleva de nuevo a un problema que ya se mencionó en el primer capítulo: el asunto de la evaluación de las preferencias para que éstas sean preferencias cualificadas o “consideradas”. En la última pintura que nuestro autor hace de un yo-presente preocupado por sus futuros yoes aún hace falta que se muestre cómo, en virtud de qué, o sobre la base de qué, es que un agente decide cambiar de preferencias; o cómo, o gracias a qué recursos es que las somete a crítica; entre otras razones –y me permito repetir lo que arriba comenté–, porque parece que Gauthier evita tener que hablar de referentes normativos que permitan llevar a cabo tal crítica o evaluación, sobre todo, si se trata de referentes irreducibles a preferencias. A esta dificultad habría que añadir que, en caso de que ocurra tal cambio de preferencias a raíz de un proceso de crítica, sin embargo, dicho cambio no involucra asimismo una crítica de los yoes *pasados* ni, por tanto, una reconstrucción –también crítica– por parte del agente, de la *historia* de esos yoes. Creo que cualquier lector, al encontrarse con la acertada referencia de Gauthier a los yoes *futuros*, sin embargo echaría de menos que a ésta se una el examen crítico de las preferencias de los yoes *pasados*, por parte de un agente que tenga esa capacidad reflexiva que Gauthier le atribuye a su individuo liberal.

Creo que en este tema de los yoes futuros encontramos una novedad bastante alentadora en lo que respecta al modelo de agencia que, hasta esta última parte de su texto, puede advertirse en Gauthier. Pues esta preocupación que ahora nuestro autor le atribuye a su agente –por las preferencias de sus yoes futuros– implicaría pensar un sujeto capaz de comprometerse con ciertas líneas de conducta y de elección, así como con ciertas normas que restringirían su comportamiento maximizador. Sin embargo pienso que habría que señalar que se echaría en falta, repito, la inclusión de las preferencias de los yoes *pasados*, las cuales también conformarían un acervo importante, y acaso indesligable de los yoes futuros, en tanto que el pasado vivido por el agente ha sido objeto de su conocimiento y de una experiencia que, si se está pensando

en un agente crítico como el ahora intenta esbozar Gauthier, creo que también tendría que ser objeto de examen por parte de éste. Así, el sujeto habría de establecer, mediante dicha crítica, en qué se identifica y en qué no con sus yoes pasados o con sus anteriores preferencias y, en estrecha relación con esto, tendría que poder reflexionar sobre *el porqué* de tal identificación o de tal rechazo y, por ende, sobre las *razones* y los *referentes normativos* a partir de los cuales decide rechazar o aprobar esas preferencias anteriores. Tales referentes acaso sean importantes para idear planes de vida y, por tanto, para proyectar sus yoes futuros; para identificar aquello con lo cual quisiera comprometerlos, teniendo en cuenta la historia del sujeto, tal y como éste la evalúe.

Pienso que una descripción como la que aquí sugiero y que, amén de involucrar los planes para los yoes *futuros*, también incluiría la evaluación de los yoes *pasados*, podría garantizar mejor esa conexión más “fuerte” que Gauthier busca ahora establecer entre el yo presente y sus yoes futuros. De modo que al yo actual se lo pueda concebir más claramente como *reflexivo* y, por tanto, como *autónomo* y como poseedor de una capacidad para *evaluar sus preferencias*, lo cual conllevaría una noción más sustantiva del agente moral. No obstante, esta es la noción que, precisamente, creo que no logra hacerse del todo clara en la descripción que nuestro autor hace, dado que en ella no se incluyen referentes normativos, ni un examen de los yoes pasados a partir de dichos referentes ni, por tanto, una base para hacer juicios de valor sobre el tipo de yoes futuros con los que quisiera comprometerse el yo actual. De allí que también, a mi entender, el lector de Gauthier extrañaría buena parte de la sustancia que uno podría esperar de una noción más “fuerte” de la unidad y de la identidad del yo, así como de la importancia que dicho yo le concedería a su autonomía y a su capacidad para autocriticarse y criticar a otros.

En los siguientes dos capítulos intentaré mostrar el contraste entre una versión del agente tal como la que acá nos presenta Gauthier, y otras versiones en las cuales creo que pueden hallarse ciertos elementos que permitirían una unidad, en mi opinión, más “fuerte” del yo, unidad que aparece como tema de preocupación de nuestro autor hacia el final de su texto. Para mostrar el mencionado contraste, me valdré de los aportes de dos autores que pienso que presentan una explicación más clara acerca de *cómo* podría operar esa capacidad crítica y autocrítica del yo que, yendo más lejos que Gauthier, no se ejerce únicamente como una evaluación de los yoes *futuros* por parte del

yo *presente*. Así, en los *yoes-pasados* el agente también puede reconocerse o, por el contrario, puede renegar de ellos pero, en todo caso, esto se daría mediante un proceso en el cual necesariamente tendría que apelarse a *razones* y a *referentes normativos*. Los cuales, en primer término, no colapsarían en *meras preferencias*; en segundo lugar, escaparían al mero autointerés o al cálculo maximizador y, finalmente, permitirían que el yo se entienda y se critique a sí mismo insertándose en un continuo narrativo o histórico, que también sería susceptible de crítica y de reconstrucción.

La reconstrucción y narración de esa historia, la cual constituye una reconstrucción *de sentido*, creo que permitiría que se pudiera ver más claramente al yo como ese continuo sustantivo al que Gauthier parece querer apuntar en esta parte final de su libro, ya que, a mi entender, una auténtica reconstrucción-narración de la propia historia, así como un imaginar y examinar *yoes-futuros*, se harían imposibles, insisto, a menos que se apele a las razones y a los referentes normativos necesarios para adelantar una reflexión crítica sobre sí mismo (e incluso sobre otros agentes). Esta reflexión que, repito, además de los *yoes futuros*, también tendría por objeto a los *yoes pasados* –en su diálogo con el yo actual–, y que tal vez se echaría en falta en la descripción de Gauthier, creo que apoyaría una idea más plausible de la autonomía y de la conciencia crítica que ese yo tenga de sí mismo y de las etapas anteriores o posteriores de su yo actual, así como de su juicio sobre sus relaciones con otros agentes. De este modo, si a esta interesante idea que nuestro autor nos presenta de un yo preocupado por sus *yoes futuros*, se añade una conexión con referentes normativos y con los *yoes pasados*, entonces se podría ir un poco más allá de Parfit y, con ello, podría verse mejor *en virtud de qué* se le puede atribuir al agente una preocupación por las preferencias de sus *yoes futuros*,¹²⁴ preocupación que, insisto, se haría indesligable de la reflexión o evaluación crítica sobre sus *yoes pasados*.

3.4 El retrato definitivo del liberal ¿“Libre afectividad” vs. cohesión social? ¿Autonomía vs. condicionamientos sociales?

Gauthier cierra esta parte de su exposición de *La moral por acuerdo* (pp. 342-345) dando por claramente demostrada la conclusión de que su agente liberal va mucho

¹²⁴ Sobre la este tema de la previsión de las futuras preferencias, véase: Nozick (1995, capítulo V); Sen (1999, pp. 54-86); Elster (2002, pp. 37-40).

más allá del *homo oeconomicus* en cuanto a su capacidad crítica, tanto en lo que se refiere a la deliberación sobre el diseño de las instituciones sociales, como al examen de las propias preferencias. Esta conclusión a la que ha llegado el autor tiene la importante consecuencia, según él, de que nos hace ver la necesidad de arrojar nuevas luces sobre un tema crucial, anteriormente comentado: la relación de imbricación que el filósofo canadiense cree encontrar entre el agente autónomo y aquello que ha llamado una sociedad “esencialmente” justa. Esto es, una sociedad conformada por agentes “completamente” (*fully*) racionales, que contrasta con otro tipo de sociedad a la que sólo pueda considerarse como “instrumentalmente” justa, *i.e.* conformada por agentes que son racionales únicamente en el sentido de ser maximizadores preocupados por satisfacer sólo sus preferencias actuales (p. 344). Este último tipo de sociedad estaría diseñada de tal forma que sus instituciones serían aptas para satisfacer las preferencias *actuales* de unos agentes pensados como meros yoes-del-mercado y como individuos sin capacidad crítica frente a dichas preferencias.¹²⁵ La importancia de la autonomía y de su imbricación con una sociedad esencialmente justa implica también el interés que tienen los agentes en la participación en los asuntos públicos, dado que les importa que éstos marchen de tal forma que respondan a la crítica ejercida por los individuos sobre sus propias preferencias y sobre el diseño y el funcionamiento de las instituciones. Pero este interés en lo público no debe hacernos perder de vista, insiste Gauthier, que la valoración que cada agente ha hecho de las instituciones sociales es ante todo de

¹²⁵ Este segundo tipo de sociedad, según Gauthier, sería inestable frente a las variaciones que sufran las preferencias de los agentes. De manera que al cambiar éstas se desharían igualmente los contratos y acuerdos entre ellos, comenzando por el contrato social de base. Por el contrario, para el autor, una sociedad *esencialmente* justa, al ser el producto de una elección arquimédica, elección que exige que cada agente que participe en ella sea *autónomo* y, por tanto, proclive a *criticar* sus preferencias, y a tomar en consideración aquellas otras que podrá tener hacia *futuro*, es una sociedad cuyas instituciones se eligen de un modo tal que permitan incorporar esta capacidad crítica de los agentes, y los cambios que dicha capacidad pueda traer consigo. Para el filósofo canadiense esto *no sólo no* debe llevarnos a pensar que ello signifique que este tipo de sociedad esencialmente justa sea de suyo inestable, sino que, por el contrario, habría que concluir que el hecho de que a dicha sociedad siempre se la pueda examinar, situándose los agentes en el punto de vista arquimédico cuando lo consideren necesario y, desde allí, examinar sus instituciones, significa una flexibilidad que garantizaría la justicia de las instituciones, por la vía de su adaptabilidad; y por ende, la continuidad de la base fundamental de la sociedad, a pesar de los cambios que le sobrevengan. El que unos agentes así de reflexivos y autónomos puedan acceder al punto de vista arquimédico siempre que encuentren razones para ello, es lo que permite, según Gauthier, que las instituciones de una sociedad esencialmente justa sean al mismo tiempo flexibles, estables y justas. En breve veremos que, sin embargo, la estabilidad de este tipo sociedad no se debe, para nuestro autor, a que sus agentes tengan un compromiso con la moral por mor de la moral misma. Sobre la estabilidad institucional y su relación con la racionalidad de los agentes sociales, véase: Elster (2002, pp. 111-193), y North (1994).

carácter *instrumental*: las ha elegido a éstas porque parecen ser aptas para la satisfacción de *sus preferencias*. De este modo, la justicia *esencial* tiene también un carácter *instrumental*. De manera análoga, la razón tiene un carácter instrumental, en virtud del cual tanto ella, como la justicia, si bien no pueden considerarse únicamente como medios para el logro de los fines particulares, es innegable que de ambas se exige que sean los mejores medios para tales fines (p. 345).

Así, la razón es sumamente útil para un tipo de agente que tiene la capacidad y la necesidad de elegir entre dos o más acciones posibles. De allí que sea objeto de una especial valoración por parte de ese agente. Algo similar sucede con la justicia: ésta resulta ser de gran utilidad para individuos que tienen la capacidad y la necesidad de elegir entre diversas formas posibles de interacción con sus congéneres, lo cual explicaría el gran valor que los agentes humanos le conceden al hecho de que las normas sociales sean justas. Para Gauthier, valoramos tanto a la justicia porque ella aporta la base de unas formas de interacción que están libres de coerción y que son la mejor manera de que cada uno pueda satisfacer sus preferencias, independientemente de cuáles sean éstas o cuáles sean las capacidades de cada agente. De manera que si la justicia y la razón se valoran de forma tan especial es porque son fundamentales para la satisfacción de dichas *preferencias*: estas últimas son el objeto primero de preocupación por parte de quienes eligen las instituciones sociales y por ello no se las debe perder de vista como el motivo fundamental que pone en marcha el proceso para lograr un pacto social.

Sin embargo, advierte Gauthier, esta preeminencia de las preferencias particulares no debe llevarnos a darle la razón al sofista. Por tanto, para nuestro autor, sería erróneo afirmar que la justicia es sólo la segunda mejor manera de lograr satisfacer dichas preferencias, en vista de que la forma óptima –la violencia unilateral ejercida por nosotros y sin que nadie nos oponga resistencia– queda por fuera de nuestras posibilidades (p. 345). Gauthier cree que es claro que ocurre todo lo contrario: la justicia es el mejor medio para lograr nuestros fines, y lo es en razón del modo en que ella opera, esto es, sin coerción, y gracias a la estructura cooperativa o participativa en la que se sitúan los agentes que se someten –voluntariamente– a sus dictámenes, así como por los bienes que se obtienen gracias a dicha cooperación. Pues, como ya se ha dicho, sólo mediante ella hacemos un uso mucho más efectivo de nuestras capacidades,

un uso que supera al que haríamos en medio de la lucha propia de un estado de naturaleza. Al mismo tiempo, únicamente por medio suyo podemos obtener aquellos fines que estarían por fuera de nuestro alcance si nos empeñáramos en permanecer en dicho estado. Los sofistas tendrían razón cuando afirman que un ser que fuese completamente poderoso y autosuficiente no le encontraría ninguna utilidad a la interacción no violenta con otros agentes, esto es, a la justicia. Pero el hecho innegable es que nosotros, los seres humanos, no somos ni autosuficientes ni completamente poderosos; necesitamos cooperar con nuestros semejantes para poder ensanchar nuestros limitados poderes y paliar nuestra insuficiencia y menesterosidad. De forma que, para seres como nosotros, la justicia no es de ningún modo un mal necesario. Al contrario: si se insiste en presentarla como si lo fuera, esto equivaldría a presentar a la propia naturaleza humana como algo radicalmente “malo” o “defectuoso”.¹²⁶

Lo anterior debe llevarnos, según Gauthier, a la conclusión de que su propuesta de una moral por acuerdo logra cerrar la brecha entre las exigencias de la razón y las de la moral. Pues, tanto para quien hace un juicioso y realista balance de sus capacidades, restricciones y posibilidades, con el fin de satisfacer sus preferencias y deseos, esto es, para el *homo oeconomicus*, como para quien está interesado en obtener los bienes que ofrecen las empresas de *cooperación* humana, es decir para el *individuo liberal*, la moral por acuerdo posee la ventaja de que captura los *intereses* del *primer* tipo de agente, así como los *afectos y entusiasmos* del *segundo* (p. 345). La conclusión que, finalmente, cree obtener Gauthier, es que la moral por acuerdo casa perfectamente con esta figura del agente liberal, cuyos afectos, entusiasmos e intereses nuestro autor sintetiza en la siguiente descripción:

1) El individuo liberal es un ser activo que encuentra satisfacción en el esfuerzo y en la lucha que caracterizan a las actividades humanas, pero de manera que *no sólo* las valora a éstas *por sí mismas*, sino *también* por los fines o *resultados* que se obtienen mediante ellas. Así, este agente, que opera en un medio caracterizado necesariamente por la escasez, tiene una noción de vida buena (*sic.* p. 346) en la que se combinan el logro obtenido, esto es, el aspecto instrumental, con el esfuerzo o la búsqueda de dicho

¹²⁶ En lo que los sofistas sí tendrían razón, según Gauthier, es que la justicia, si es una virtud apropiada solamente para las necesidades humanas, por contra, no sería nunca algo útil para quien se viera a sí mismo como un dios o se quisiera poner por fuera de los límites humanos, o se negara a pagar los costes (y a recibir los beneficios) de quienes estamos en condiciones meramente humanas (p. 345).

logro; es decir, se armonizan el aspecto *intrínseco* de la valoración que hacen los agentes de los *bienes* que buscan y el aspecto *instrumental* de las *actividades* mediante las cuales se obtienen tales bienes.

2) El agente liberal posee su propia concepción particular de bien, la cual no necesariamente descansa en una motivación asocial, dado que se trata de un individuo que, por definición, *no* es asocial. Sin embargo, al ser tan diversos los bienes que persiguen los distintos agentes, pueden presentarse, tanto situaciones de conflicto, como de armonía o consonancia entre la búsqueda que cada cual hace de su propio bien. Este último es el producto apenas esperable o natural de la expresión de las preferencias que cada agente tiene pero, si se trata de preferencias cualificadas o “consideradas”, se pone de presente que, más allá del conflicto o de la armonía, lo importante es la capacidad que tiene cada agente para reflexionar sobre sus preferencias y cambiarlas si encuentra razones para ello. En síntesis, la figura del individuo liberal se halla caracterizada de forma esencial por el hecho de que se trata de un agente con capacidad para el ejercicio de su autonomía.

3) Al poseer esta capacidad reflexiva, este tipo de agente es “completamente” (*fully*) racional y sus competencias racionales incluyen tanto su autonomía para elegir, perseguir o *criticar* una concepción de *bien*, como su capacidad *maximizadora*, *i.e* el poder para elegir entre diferentes cursos de acción a partir de dicha concepción de bien, esto es, a partir de sus preferencias consideradas. Así, para este agente la utilidad que persigue es la medida de dichas preferencias, de tal forma que se trata de un agente *maximizador* de utilidad esperada y, por tanto, de un agente racional, si bien Gauthier insiste en que en esta figura del agente liberal también juega un papel fundamental su capacidad *crítica-reflexiva*, en contraste con la “cruda” racionalidad económica (p. 346).

4) A la mencionada capacidad para evaluar y, eventualmente, modificar sus preferencias, no se la debe pensar como algo que se ejerza de manera arbitraria, sino como una competencia que obedece a la autonomía fundamental del agente, y en virtud de la cual éste se caracteriza por tener un sentido de su “yo” en tanto que permanente a lo largo del tiempo, y a pesar de los cambios que pueda sufrir. Posee, por tanto, una identidad; se sostiene como una unidad que se integra, que no se disgrega, por más variaciones que puedan presentarse en su contexto o en sí misma. Es así como el yo

puede percibir su presente en tanto que conectado con sus futuros yoes, al tiempo que su preocupación por dicho futuro implica que el agente se hace conciente de la necesidad de restringirse en su conducta maximizadora. Para lo cual adopta ciertas “políticas” de comportamiento en el largo plazo, es decir, ciertos “principios”¹²⁷ de conducta o de elección con los cuales se compromete, al darse cuenta de que éstos le resultan sensatos o racionales, si los ve como restricciones justificadas a su actividad maximizadora, en tanto que dichas restricciones sean vistas como medios para evitar los riesgos del conflicto con otros agentes. Se trataría, entonces, de principios “morales” que el agente adopta voluntariamente, dado que percibe su acatamiento como conveniente en el largo plazo, en el sentido de que dicho acatamiento es visto como un medio necesario para la maximización de su utilidad esperada y, por ende, para la satisfacción de sus intereses y preferencias.

5) Sin embargo, esta capacidad de previsión y la prudente obediencia a principios morales no son sólo una expresión del *cálculo racional* propio de un astuto maximizador, sino también la expresión de una capacidad propia del individuo liberal y que, para Gauthier, significa un *plus* muy importante frente al *homo oeconomicus*: la *capacidad afectiva* o emotiva para comprometerse con lo moral. Esto guarda una estrecha relación con el hecho de que el agente liberal se da cuenta que puede elegir entre diversas formas de vida y que la riqueza de éstas depende de su interacción con *otras personas*, las cuales, a su vez, bien pueden elegir estilos de vida distintos al suyo. Es decir, el agente liberal es conciente de que su vida sólo puede transcurrir en medio de un contexto social, en medio de las interacciones con otros agentes, al tiempo que valora positivamente estas oportunidades de enriquecer su vida a partir del contacto con esas otras personas. De allí que también valore de manera especial la participación con ellas, tanto como valora su propia actividad y sus propios esfuerzos.

6) No obstante, hay que tener siempre presente que el hábitat del agente liberal está signado, no sólo por la presencia de otros agentes —el factor social— sino también por la condición ineludible de toda vida humana: la escasez. De allí que la satisfacción *intrínseca* que el agente liberal obtiene del primer factor, es decir, de sus *interacciones sociales*, es indesligable del rol *instrumental* que éstas cumplen en la solución de los

¹²⁷ Sobre la racionalidad de la adopción de principios, véase: Nozick (1995, capítulo I); Sen (1999, pp. 54-86).

problemas que acarrea el segundo factor, la *escasez*. Esto implica que en el agente liberal se entrelazan la *racionalidad*, orientada a la solución de dichos problemas, esto es, a la maximización de su propio interés, y la *afectividad* nacida de la necesidad de interactuar con otros agentes, y de las satisfacciones que dicha interacción le depare. “Valuing participation and her fellow participators, she values also the constraints that make it possible; both intellect and emotion make her a moral being” (p. 347).

7) Pero esta afectividad, si bien obedece en último término a una necesidad humana, dictada por nuestras condiciones de escasez y de menesterosidad, no obstante, en el caso del agente liberal, es ejercida libremente, sin coacción. El individuo es libre de elegir a quiénes va amar: sus lazos de afecto y de lealtad no le vienen impuestos externamente ni por nexos de familia ni de tribu. De manera análoga, ni las actividades a las que quiera dedicarse, ni sus roles sociales le son impuestos; por lo tanto, el agente no se encuentra atado de una manera fija ni de por vida a dichos roles y actividades.

Esta “libre afectividad” de los individuos liberales es, según Gauthier, consistente con la justicia *esencial* que caracteriza a la sociedad que ellos conforman, la cual por definición debe consistir en una empresa de cooperación orientada al *beneficio mutuo* y, por tanto, ha de permitir que, partiendo de lo que cada cual considera bueno para ella/él, cada uno elija con quiénes y de qué maneras quiere asociarse. El autor advierte que algunos podrían temer que esta libertad afectiva (*free affectivity*) se traduzca en una peligrosa disolución de los lazos sociales. Sin embargo, Gauthier cree que, por el contrario, dicha forma de libertad es la clave para el verdadero cemento de la sociedad,¹²⁸ ya que éste sólo podría darse si las relaciones entre las personas son percibidas por ellas como algo que han elegido voluntariamente y en virtud de sus

¹²⁸ Según Gauthier, tal efecto de cohesión social logrado a partir de este tipo de libertad individual, obedecería a dos razones. En *primer* lugar, si la sociedad es vista por sus miembros como un auténtico sistema de cooperación para el beneficio mutuo, entonces en ella la voluntad de interactuar con otros agentes va depender mucho menos, como ya se mencionó, de que exista un previo lazo de familia o de clan. En *segundo* término, sólo en una sociedad en la que esto ocurra, puede esperarse que florezca una auténtica “amistad cívica” (*sic.* p. 348), pues ésta sólo puede surgir allí donde a las personas *no* se les imponga por la fuerza ni su lealtad con una idea de bien social, ni el cumplimiento de unos roles sociales fijos en razón de dicha idea. De allí que la gente que puede elegir libremente sus lazos de afecto y sus lealtades, así como sus actividades y roles, puede también, según Gauthier, responsabilizarse mejor por ellos y hacer honor a los compromisos que adquiere en razón de tales afectos y roles sociales. El autor reconoce que su propuesta de una “libre afectividad” posiblemente pueda preocupar a quienes se empeñen, según él, en defender la idea de una “comunidad sustantiva”, así como también a aquellos que teman que con la libre afectividad se pueda generar insolidaridad social e intergeneracional, si bien nuestro autor no descarta que esto último pueda presentarse en algunas sociedades liberales (pp. 348-349). Por cierto, con respecto al tema de la amistad, Gauthier no menciona a Aristóteles.

gustos e intereses; y no como una imposición externa que obedece a jerarquías, identidades o roles sociales fijos y forzosos. De allí que su descripción del agente liberal debe mostrarnos, según el filósofo canadiense, la importancia de una defensa del individuo y la prioridad que éste debe tener sobre la comunidad. Esto *no* se relaciona, al decir de Gauthier, y como ya lo había mencionado anteriormente, con ninguna tesis acerca del *origen* de la conciencia individual. El autor dice estar de acuerdo en que muy seguramente dicha conciencia se ha generado como un producto del proceso de socialización, lo cual, sin embargo, no debe llevarnos a renunciar a la idea de la autonomía ni al modelo de agente como individuo liberal ni, menos aun, debe llevarnos a darle la razón a quienes piensan que la moral por acuerdo no es más que la pseudomoral del *homo oeconomicus*. Creo que en este punto tal vez Gauthier se percate de la crítica que se le puede hacer a su modelo de agente en tanto que podría parecer que nuestro autor lo presenta como un ser que se autoconstituye independientemente de todo contexto social; como alguien que posee desde siempre plena capacidad de agencia y de autonomía, sin que en ello hayan intervenido condiciones sociales que le hayan precedido y determinado. Ante críticas de este tenor, el canadiense responde, en primer término, y como ya lo ha repetido antes, que su propuesta no está orientada a explicar la génesis de la conciencia individual ni acarrea consecuencias para esta cuestión y, en segundo lugar, también reconoce que es indudable que la autonomía requiere de unos “materiales” o “recursos” fácticos previos a partir de los cuales pueda ejercerse. Esto es, la crítica o la reflexión que lleva a cabo el agente sobre sus propias preferencias y sobre sus intereses necesita, precisamente, que haya unas preferencias y unas capacidades previas que permitan llevar a cabo tal evaluación.

Estas capacidades y preferencias serían, entonces, los “inputs” de tal proceso de evaluación, inputs que han surgido a partir de las condiciones en las que se encuentra el agente, no siendo, por tanto, un producto de su autonomía, sino, por el contrario, una condición previa al ejercicio de ésta. Sin embargo, esto no debe hacernos olvidar, advierte Gauthier, que lo que hace que un individuo sea autónomo no es que tenga preferencias ya formadas, sino su capacidad para *cambiar dichas preferencias* mediante un proceso de *reflexión*. Este último sólo puede comenzar en individuos que ya han sido socializados y cuyas preferencias y capacidades, por lo tanto, estarían en cierto modo socialmente determinadas, de la misma manera en que la capacidad crítica y reflexiva

también, en parte, requeriría de unas condiciones previas. Éstas, así mismo, constituyen el terreno en el que tiene lugar ese proceso de socialización que ha marcado al individuo y que igualmente determina su capacidad crítica (p. 349), sin embargo, lo importante para destacar de este proceso de socialización, según Gauthier, es que no constituye a las personas a la manera en que el *hardware* de un ordenador constituye al sistema. Más bien, lo hace a la manera de un *software* (p. 350), razón por la cual no debe en absoluto desecharse la idea de que siempre está abierta la posibilidad de ejercer la autonomía, si bien hay que reconocer la necesidad de unas condiciones para dicho ejercicio. Para Gauthier, esto último también permite caer en cuenta de que siempre es posible una sociedad *esencialmente* justa, conformada por individuos *liberales*, en quienes, según el filósofo canadiense, cualquier lector puede ver ahora, luego de esta última parte de su argumentación, una forma de agencia cuyas capacidades irían mucho más allá que aquellas que caracterizan al mero *homo oeconomicus*.

4. Conclusiones parciales

En estos tres primeros capítulos se ha podido observar una suerte de evolución en el retrato que Gauthier va pintando de su agente modélico. Se ha partido de la figura de un maximizador egoísta pero también prudente (MR) que está dispuesto a restringir en parte la búsqueda de su propio beneficio, cumpliendo para ello con ciertos constreñimientos morales, con los cuales se compromete por mor de la maximización de su utilidad personal. La motivación moral de este tipo agente se reduce, pues, al autointerés, al tiempo que los lazos que este individuo teje con otros agentes no son otros que aquellos que resultan de las relaciones contractualmente mediadas. De allí también que el escenario en el que se construyen tales relaciones sea el mercado y que el agente modélico que inicialmente nos presenta el autor sea, precisamente, el *homo oeconomicus*. De esta figura queda excluido todo elemento que implique una base motivacional más amplia, v.g. el compromiso con ciertos valores o con ciertas normas a las que se busque hacer honor por mor de ellas mismas, el altruismo, la preocupación por otros agentes, etc. Igualmente, también se excluye toda forma de deliberación que vaya más allá del cálculo estratégico sobre la viabilidad de los medios necesarios para lograr los fines que el agente egoísta y maximizador se proponga. De este modo, el

agente no se detendría a evaluar sus preferencias, salvo para asegurarse de que éstas guarden entre sí relaciones de consistencia, y se apoyen en una base epistemológica confiable. De manera que este yo-del-mercado sería reflexivo en un sentido muy reducido, del mismo modo en que puede calificarse de reducidos sus recursos deliberativos, así como el repertorio de sus experiencias moralmente relevantes.

La motivación que tenga este agente para la moralidad dependería de lo que éste perciba en otros agentes, de modo que no tendría incentivos para cumplir con normas morales, a menos que crea que los demás individuos también lo van a hacer. Su desinterés en la suerte de estos otros agentes le convierten en un ser desvinculado, un Robinson que, a expensas de la pobreza de sus lazos con otros individuos, logra una forma de libertad que, para Gauthier, resulta bastante atractiva. Dicha libertad no podría lograrse sino en el contexto de un mercado en el que el cumplimiento de las normas morales asegure la no explotación y el mutuo beneficio para todos los que en él participan. Estos Robinsones pueden, por consiguiente, ser pensados como seres abstractos, en el sentido de que, a causa de su desinterés mutuo, parecieran carecer de un contexto social, cultural e histórico, esto es, un contexto en el cual se hayan construido sus relaciones interpersonales. Cada uno de ellos es, pues, un ser auto-creado, que para ejercer su capacidad de agencia no ha necesitado de otros individuos, que ha gozado de dicha capacidad desde siempre y, por tanto, se trata de alguien que no tiene una biografía, un historia vital que pueda narrar, en la que hayan jugado algún papel otros agentes con los que haya podido tejer vínculos sociales, y en la cual se haya forjado algún tipo de identidad. Por ende, se trataría de un ser sin cultura, sin lenguaje y sin los compromisos valorativos que tales vínculos hubieran podido inculcar en su repertorio posible de referentes de significación o de sentido. De manera que este vacío bien podría ser considerado como el causante de que este tipo de agente carezca de elementos normativos que le permitan evaluar sus preferencias o que le hagan competente para enfrentarse a un dilema moral.

Si a esto se agrega que su motivación para cumplir con las normas morales no va más allá de su autointerés, el resultado implica una cierta inestabilidad en el compromiso con tales normas y, por ende, el peligro de que el pacto social que suscriban agentes a los que se piense de este modo, se vea traicionado en caso de que surjan incentivos para hacerlo. De allí que Gauthier necesite retocar y matizar la figura

de su agente maximizador egoísta, para que éste, así como el escenario en el que se relaciona con otros agentes, se “pacifiquen” y se enriquezcan, y comiencen a incluir elementos tales como el gusto por la socialización y diversas formas de interacción que vayan más allá de las relaciones económicas. Por ello, en un segundo paso en el proceso de transformación de su agente modélico, Gauthier incluye en el retrato de este personaje ciertas novedades, por ejemplo, aquello a lo que nuestro autor se refiere como “motivaciones secundarias”, las cuales apuntan a un interés por la sociabilidad. No obstante, Gauthier insiste en que éstas se superponen a las motivaciones “primarias”, que son más fundamentales o que son la base de las primeras, y que vienen determinadas por el autointerés y la necesidad de desplegar una conducta maximizadora. Dicho autointerés y dicha conducta siguen, entonces, considerándose como fundamentales e ineludibles, dadas las también insoslayables condiciones humanas de la escasez de recursos y de la búsqueda de la autoconservación. Nos hallamos, pues, ante un agente que, frente a su primera versión, presenta el plus de que aprecia a sus semejantes, así como el enriquecimiento personal —no sólo económico— que obtiene de su relación con ellos, pero que no ha dejado de ser un maximizador autointeresado, si bien ya no se trata del mero *homo oeconomicus*.

La culminación de este proceso de transformación del agente modélico de Gauthier es el retrato del individuo liberal que, como acabamos de ver en este tercer capítulo, presenta una base motivacional “mixta”, en el sentido de incluir afectos y entusiasmos relacionados con su gusto por la sociabilidad, al tiempo que sigue siendo un sujeto autointeresado en el que, y sin que el propio agente sea consciente de ello, dicho gusto tiene un valor instrumental frente al fin más importante que es la autoconservación del propio individuo. Lo que el agente ignora es que su aprecio por la sociabilidad obedece al importante papel que ésta ha jugado en la evolución de la especie humana. Dicha explicación evolucionaria y, por ende, el valor meramente instrumental de lo social, sólo son conocidos, por supuesto, por el filósofo, en este caso, por Gauthier quien, para dar cuenta de los fenómenos morales se sitúa “por fuera” de la praxis moral misma, objetivándola y no asumiendo el papel de alguien que participe en dicha praxis.

El agente liberal asimismo puede ser considerado como un ser más reflexivo que su antecesor, el *homo oeconomicus*, lo cual le permitiría evaluar más cualificadamente

sus preferencias, hasta el punto de llegar a preocuparse por aquellas preferencias que el agente vaticina que va a tener en el futuro. Es, pues, un individuo con mayor capacidad crítica y, por ende, con mayor autonomía, lo cual incluye un mayor sentido de su propia unidad como un “yo” a través del tiempo. Sin embargo, en este punto también nos encontramos con el obstáculo que representa el hecho de que el agente siga sin contar con referentes normativos a partir de los cuales evaluar sus preferencias. Por ende, también se trata de un individuo que no comparte estos posibles referentes con otros agentes y que, en general, tanto en lo que atañe a la verdadera naturaleza de su motivación para relacionarse con esos otros, como a los demás aspectos que puedan presentar las razones que el agente crea tener para su actuar, resulta inaccesible para sí mismo, y en cambio, dicha naturaleza y la auténtica explicación de tales razones, insisto, sólo son accesibles “desde fuera” para el filósofo quien, repito, es quien sí sabe que tal motivación obedece finalmente a una necesidad evolucionariamente dirigida, que no tiene otra explicación que la búsqueda de la utilidad individual.

Con miras a alcanzar una mayor claridad sobre las falencias y vacíos que hayan podido ser señalados en la versión que Gauthier presenta de la agencia humana, así como con la intención de sugerir posibles soluciones a algunos de los problemas que se han identificado en dicha versión, en los siguientes dos capítulos me ocuparé de otros dos modelos de la agencia humana, representados en los aportes de Harry Frankfurt y Charles Taylor. Intentaré que desde estos aportes se puedan obtener elementos que permitan esbozar un modelo más plausible y menos restringido de las posibilidades de la agencia humana. Estos elementos que permitirían obtener un relato más rico de las formas de deliberación, de los recursos que permitirían adelantar tal deliberación, así como las razones para actuar y la base motivacional que cabría atribuir a un agente al que tal vez podamos considerar menos abstracto, menos artificioso y más cercano a nuestras intuiciones sobre ciertas experiencias morales a las que acaso no quisiéramos considerar como ajenas al ejercicio de nuestra agencia humana, ni al modo en que quisiéramos auto-representarnos como tales agentes. Igualmente, intentaré mostrar que el examen de dichas experiencias requiere, con respecto a la actitud de Gauthier como filósofo moral, un cambio de actitud: asumir, tal y como creo que aconsejarían P.

Strawson y J. Habermas,¹²⁹ el papel de quien sí toma parte en las interacciones humanas y no se limita a objetivarlas.

¹²⁹ Cf. Strawson (1995) y Habermas (1994a).

Segunda Parte

Hacia un modelo más complejo de agencia: de la “persona” de Frankfurt al “evaluador fuerte” de Taylor

En esta segunda parte de la tesis me propongo analizar los aportes de Harry Frankfurt y Charles Taylor, con el objeto de mostrar de qué manera pueden hallarse en las propuestas de ambos autores elementos de un modelo más complejo de agencia humana, un modelo en el que se verían superadas las limitaciones que, en lo que va del presente trabajo, se han logrado señalar en el discurso de Gauthier. Me ocuparé en primer lugar de los aportes de Frankfurt para luego terminar con los de Taylor, intentando que este orden expositivo haga justicia a lo que, en mi opinión, admite ser visto como un proceso en el que la pintura del agente humano va ganando en complejidad y, por ende, va incluyendo elementos que significan una novedad frente aquellos que hayan podido identificarse en los primeros tres capítulos de la tesis. De este modo creo que el esquema de desarrollo le haría más fácil al lector advertir las similitudes y continuidades, pero también los “saltos” que creo pueden establecerse entre Gauthier y Frankfurt, de una parte, y entre este último y Taylor, de la otra. Lo cual permitiría apreciar hasta qué punto la postura de Frankfurt puede ser considerada como una suerte de “transición” entre una idea reducida de la agencia humana, tal y como la que se ejemplifica en Gauthier, y una pintura más compleja: aquella que creo cabe obtenerse de la obra de Taylor.

En el capítulo dedicado a Frankfurt en *primer* lugar pretendo mostrar la *novedad* representada por algunas de sus tesis, las cuales pienso que significan una ruptura con buena parte de los supuestos asumidos por Gauthier y por las tradiciones de las que este último se considera heredero. No obstante, en *segundo* término también aspiro a que en Frankfurt se haga patente la persistencia de ciertos compromisos que indican una suerte de *continuidad*, de remanentes o de ataduras con una visión restringida de la agencia humana. En lo que se refiere al *primer* aspecto de su propuesta, intento poner de presente la ruptura que significan, frente al modelo de Gauthier, las nociones de “persona” y de evaluación de los deseos, tal y como son desarrolladas por Frankfurt. Así, buscaré resaltar el salto cualitativo que creo que habría, por un lado, entre la

“persona” de Frankfurt y el maximizador autointeresado de Gauthier y, de otro lado, pero en estrecha relación con ello, la novedad representada por la evaluación de los deseos y el “preocuparse” (*caring*) que Frankfurt le atribuye a la persona, frente a las “preferencias” que explican las acciones del maximizador de Gauthier, así como la preeminencia que en este autor tiene el uso meramente instrumental de la razón. Intentaré mostrar que la idea de Frankfurt de que una persona está en capacidad de llevar a cabo una evaluación de sus deseos, indica una mayor complejidad, tanto en el concepto mismo de *deseo/preferencia*, como en la propia noción de un *agente* que es capaz de criticar sus propios deseos. Así mismo, me propongo mostrar la importancia que reviste dicha complejidad –que en palabras de Frankfurt es de naturaleza “volitiva”–, para efectos de una ampliación de las posibilidades de concebir la base motivacional del agente, amplitud que, a su vez, implicaría una contravención del compromiso reduccionista de Gauthier. Si se examina el importante papel unificador y, sobre todo, dador de sentido, que cumplen aquellos elementos a los que Frankfurt se refiere como las “preocupaciones” de una persona, se podrá apreciar que éstas se resistirían a ser reducidas a una base motivacional como el “autointerés” en el que tanto insiste Gauthier.

Empero, como ya se ha advertido, también me propongo que salgan a la luz algunos supuestos que siguen atando el discurso de Frankfurt al paradigma del sujeto solitario y pensado *in abstracto*, si bien dichas ataduras acaso no se vean del todo claras hasta que no se arribe al capítulo dedicado a Taylor. En dicho capítulo espero mostrar de qué modo en este último autor puede hallarse, al menos en parte, la solución al “lastre” que significan ciertos compromisos de Frankfurt con la figura de una suerte de “Robinson ” parecido al que Gauthier postula como su agente ideal. Este lastre se traduce en la dificultad que entraña el hecho de que Frankfurt insista en establecer algunas prohibiciones o “vetos” a la posibilidad de desarrollar ciertas experiencias en el ejercicio de la agencia, experiencias que, no obstante, pienso que no tendrían por qué ser objeto de dicha prohibición si se toman en serio los avances que aporta su modelo de “persona” (y de autoevaluación) frente a otras nociones más restringidas de agencia. Dentro de los mencionados vetos, se verá que resultan particularmente disonantes aquellos que implican trabas para la articulación de las *razones* que tiene un agente para actuar. Se trataría, pues, de auténticos escollos para que en el repertorio de experiencias

propias de la “persona” de Frankfurt también se incluyan elementos cognitivos que, de nuevo, si se asumen en su radicalidad los aportes más novedosos de este autor, creo que tendrían que concebirse como necesariamente imbricados en aquellos compromisos “volitivos” asumidos por el agente. En otros términos, tales compromisos no tendrían por qué ser reducidos a un ámbito puramente volitivo del agente humano, ámbito al que Frankfurt insiste en considerar como completamente ajeno a lo cognitivo, a lo ético y a lo moral.¹³⁰ En este orden, quisiera hacer ver que lo volitivo tampoco tendría por qué excluir la experiencia del dar, escuchar y examinar razones en el contexto de un diálogo que la persona tendría que poder sostener con *otros* agentes –físicos o imaginados, pero en todo caso “otros”–, y unos otros que son insoslayables.

Finalmente, en el quinto capítulo, se intentará ver de qué manera los aportes de Frankfurt pueden incluirse dentro de un modelo más amplio de agencia, en el que la mayor complejidad de la experiencia reflexiva del agente, ya dibujada por Frankfurt, sería objeto de un desarrollo más radical si se asume la postura de Taylor. Así, la evaluación de los deseos y el “preocuparse”, propios de la “persona” de Frankfurt, creo que admitirían ser vistos como parte de las experiencias –comparativamente más ricas– del “evaluador fuerte” de Taylor, en tanto que este último personaje puede adelantar su proceso reflexivo y la aclaración de sus compromisos valorativos en el contexto de un diálogo con sus “otros significativos”. Esto es, con un auditorio con el cual se comparten, se aclaran, se debaten, se critican y negocian los *referentes normativos* a los que apela el agente en sus deliberaciones, referentes que, por su misma naturaleza intersubjetiva, hacen estallar las ataduras con el maximizador autointeresado ejemplificado por el modelo de Gauthier. No obstante lo anterior, al finalizar el quinto capítulo, me permitiré señalar algunos límites con los que puede encontrarse el modelo del evaluador fuerte de Taylor, concretamente, el hecho de que la posibilidad misma de ejercer una agencia que cuente con las ventajas atribuibles a dicho personaje, no siempre ni para todos está asegurada como el destino feliz e ineludible de cualquier ser humano. Pues La posibilidad de ser críticos frente al propio “horizonte valorativo” al que se refiere Taylor, horizonte indesligable de los modelos que los agentes han tomado de su contexto social y, de este modo, la posibilidad de adelantar un proceso de

¹³⁰ Remito al lector a la explicación que, en un pie de página de la introducción de la tesis, se hace del modo en que se usarán los términos “ética” y “moral”.

autoaclaración y articulación de sus compromisos valorativos, también pasa por los avatares de la buena o mala fortuna moral que favorezca o que, por el contrario, frustre el despliegue de su autonomía, de sus posibilidades de interlocución social y de presencia política.

Habría entonces que tener en cuenta las dificultades impuestas tanto por el peso de aquello que el propio Taylor llama “imaginarios sociales”, como por las relaciones de reconocimiento o no-reconocimiento que el agente guarde con sus “otros”, en cuanto a las posibilidades y en cuanto a los límites que dicho peso impondría en la capacidad y en las oportunidades que tenga el agente de hacer concientes y de someter a crítica tales relaciones, así como de aclarar y evaluar ese horizonte de valores compartido con esos otros. En todo ello, la buena o la mala fortuna moral jugaría un papel importante en la forja de la propia autoimagen del sujeto, en sus posibilidades de autoevaluación y en la conciencia que pueda tener de su entorno social. De tal manera que, si bien ser un evaluador fuerte admite ser visto como un *telos* deseable y normativamente defendible y necesario, no obstante habría que estar atentos a los peligros que harían que dicho *telos* no siempre esté garantizado para todos los agentes; razón por la cual más bien tendría que ser pensado como un logro, una ganancia a veces duramente obtenida.

Capítulo 4

Frankfurt: la persona, su estructura volitiva y sus “preocupaciones”

En este capítulo se examinarán las implicaciones de la tesis de Frankfurt de que los seres humanos nos distinguimos de otro tipo de agentes no por nuestras habilidades racionales, sino por una compleja estructura volitiva que se manifiesta en nuestra exclusiva capacidad para evaluar la deseabilidad de nuestros deseos. Según el autor, la dinámica mediante la cual se lleva a cabo dicha evaluación demuestra que a los individuos humanos habría que pensarlos como agentes “integrados”, lo cual para Frankfurt establece un claro contraste con la noción de agente que le atribuye a la moderna teoría de la acción, pero que, a mi parecer, también se hace evidente en la propuesta de Gauthier, tal y como se la describió en los anteriores capítulos. Si bien Frankfurt sólo se refiere globalmente a la teoría de la acción, sin dar nombres de autores específicos y, por ende, aunque también es cierto que no menciona a Gauthier, pienso que puede advertirse un franco distanciamiento entre el agente “integrado” propuesto por Frankfurt y el maximizador autointeresado que surge en discursos como el de Gauthier. Para el primero, lo que nos caracteriza como seres humanos viene determinado por una necesidad nuestra de ejercer una agencia “unificada” y de comprometernos con una identidad. Dicha identidad, a su vez, se establece o se perfila a partir de aquello a lo que Frankfurt se refiere como “preocupaciones” fundamentales (*caring*) de los agentes humanos.¹³¹

Como se irá viendo a lo largo de este capítulo, la naturaleza de dichas preocupaciones y el papel que éstas cumplen para el especial ejercicio de nuestra agencia y para la explicación de por qué hacemos lo que hacemos, en mi opinión permite que se “abra” o se “rompa” esa suerte de “coraza” que cubre la base

¹³¹ Me he atrevido a añadir el adjetivo “fundamentales”, que Frankfurt no utiliza, pues creo que la sola palabra castellana “preocupaciones” (como traducción de *caring*), si no va acompañada de este adjetivo, no captura del todo el énfasis que pienso que le interesa hacer a Frankfurt cuando se refiere al *caring* que le atribuye a aquel tipo de agentes a los que se refiere como “personas”. Es cierto que he podido utilizar los vocablos castellanos “cura” o “preocupación” (en singular), este último sin el adjetivo “fundamentales”, pero me temo que con ello el lector no se haría fácilmente cargo del énfasis que considero importante en Frankfurt. En lo que sigue, confío en que se vea la justificación que cabe a esta licencia que aquí me he permitido.

motivacional que, en posturas como las de Gauthier, es asumida como incuestionable o inanalizable, so pena de caer en honduras metafísicas atribuibles, como vimos en dicho autor, a compromisos de carácter religioso o moralista, ajenos al análisis filosófico. Me refiero a la base motivacional que Gauthier resume en su fórmula del “autointerés”, el cual, como se vio, está firmemente conectado su tesis de que las capacidades deliberativas que podemos desplegar los seres humanos, si bien pueden llegar al límite de que logremos “cualificar” algunas de nuestras preferencias –aunque sin apelar para ello a referentes normativos irreducibles a preferencias– siguen siendo, no obstante, las capacidades propias de un maximizador ocupado en determinar únicamente los medios más idóneos para la obtención de sus fines, unos fines que están por fuera del alcance de cualquier evaluación “racional”.

En Frankfurt, como veremos, si bien se conserva buena parte de esta opacidad de los fines personales, no obstante creo que habría una diferencia muy importante frente a un punto de vista como el de Gauthier; a saber, que no solamente podemos idear planes de acción para lograr lo que queremos y satisfacer nuestros deseos, sino que también podemos *evaluar* esos mismos deseos. Por esta misma vía, también podemos ser seres que se autoevalúan y, por consiguiente, seres cuyas motivaciones abarcan un espectro mucho más amplio que las “preferencias” a las que se refiere Gauthier, por más “cualificadas” que éstas sean para el autor canadiense. Sin embargo, en este capítulo también irán saliendo a la luz algunos de los problemas que surgen dado el esquema jerárquico según el cual Frankfurt explica la evaluación de los deseos, y la naturaleza que el autor le atribuye a las preocupaciones que vertebran la vida volitiva de las personas, ya que, como arriba se ha advertido, Frankfurt insiste en excluir de dicha naturaleza elementos cognitivos que le permitirían al agente apelar a ciertos referentes normativos en el momento en que el individuo se vea impelido a dar cuenta de sus preocupaciones o de aquellas decisiones que tome teniendo en mente a estas últimas.

Así, a pesar la novedad que representan las tesis de Frankfurt frente a posturas como las de Gauthier, aun teniendo en cuenta que con dichas tesis parece abrirse una puerta prometedora para una explicación más convincente de la complejidad de la agencia humana, se llega a un tope que, sin embargo, trunca la posibilidad de que el agente pueda *dar razones* de su acción de cara a *otros* agentes. Creo que sólo si se incluyera este elemento *cognitivo* se obtendría un cuadro más claro y convincente del

importante papel que, en mi opinión, tan lucidamente le atribuye Frankfurt a las preocupaciones como “filtro” de los deseos del agente y como vértice de sus elecciones, y de aquello que este autor llama “identidad”. Así mismo, podría hacerse menos artificioso su modelo de una “jerarquía” de deseos y metadeseos, modelo que, como veremos, ha sido duramente criticado por algunos autores que ven, tanto en él, como en las nociones conexas de “identificación” e “incondicionalidad” (*wholeheartness*), el peligro de caer en los abismos de un emotivismo/irracionalismo que signifique un retroceso hacia la figura de un agente carente de recursos reflexivos y argumentativos para dotar de sentido sus compromisos y elecciones.

1. Complejidad volitiva y deseos “de segundo orden”

En su artículo “Freedom of the Will and the Concept of a Person” (1998, pp. 11-15),¹³² Frankfurt se pregunta qué es lo que distingue el particular modo de ser-agentes de ese tipo de seres a los que llamamos “personas”. Si se repara en el hecho, innegable según el autor, de que hay seres de otras especies con los cuales los humanos compartimos no sólo la capacidad básica de desear y de actuar movidos por esos deseos, sino también las capacidades racionales que se requieren para idear, elegir y efectuar planes de acción orientados a la satisfacción de tales deseos, nuestro filósofo concluye que es un error insistir en el trajinado tópico de que aquello que nos distingue como “personas” sean nuestras habilidades *racionales*. En realidad, según Frankfurt, lo que determina la particular forma de agencia de esos seres a los que llamamos “personas” es la estructura de su voluntad (*will*), estructura cuya complejidad sí determina una llamativa diferencia entre una “persona” y cualquier otro tipo de agente racional con el que se comparta la posibilidad de desplegar habilidades deliberativas y estratégicas. Frankfurt insiste en hablar de *complejidad volitiva* porque una persona, amén de desear, también puede *evaluar* sus deseos. Esto es, puede desear o no desear tener los deseos

¹³² Tomo la versión que acá comento de una compilación de artículos titulada *The Importance of what we Care about. Philosophical Essays*, la cual fue publicada en 1998, si bien dicha compilación contiene textos que datan de los años 70s y 80s. Este detalle cronográfico, como veremos, creo que puede ser importante en cuanto a la idea que podamos hacernos sobre la evolución del pensamiento de Frankfurt. En adelante, las referencias a su obra las haré dentro del cuerpo del presente capítulo, no a pie de página, señalando la fecha de publicación del artículo (o de los artículos citados) y, eventualmente, la paginación correspondiente.

que de hecho tiene, o desear tener otros deseos que no son los que efectivamente le mueven a actuar.

Los seres humanos podemos, entonces, ser reflexivos y tomar una distancia crítica frente a nuestras propias motivaciones, propósitos y deseos, de tal manera que podemos evaluar la “deseabilidad” (*desiderability*) de éstos. De allí que, a diferencia de esas otras especies con los que compartimos habilidades deliberativas y estratégicas, podamos, además, tener una subclase especial de deseos: aquellos que Frankfurt llama “deseos de segundo orden”, es decir, deseos sobre deseos. La particularidad de la estructura volitiva de un agente al que llamamos “persona” se revela, entonces, en ese salto cualitativo que se presenta entre los deseos “de primer” y los deseos “de segundo” orden; salto que se evidencia en el componente evaluativo y autoevaluativo presente en estos últimos y de manera exclusiva en los agentes humanos los cuales, por tanto, bajo esta luz pueden verse como seres con una especial capacidad reflexiva.

It is my view that one essential difference between persons and other creatures is to be found in the structure of a person's will. Human beings are not alone in having desires and motives, or in making choices. They share these things with the members of certain other species, some of whom even appear to engage in deliberation and to make decisions based upon prior thought. It seems to be peculiarly characteristic of humans, however, that they are able to form what I shall call “second order desires”. [. . .] Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are. Many animals appear to have the capacity for I shall call “first order desires” or “desires of the first order”, which are simply desires to do or not to do one thing or another. No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires (1998, p. 12).

Pienso que, como puede observarse, Frankfurt parece ir en contravía del primado que ha tenido la noción de racionalidad estratégica-instrumental¹³³ en una importante corriente de la tradición filosófica moderna, y en lo que respecta a los alcances que pueda tener una explicación de las acciones humanas que tome como incuestionable dicho primado. Habría que advertir, sin embargo, que nuestro filósofo sólo se pronuncia genéricamente contra los teóricos contemporáneos de la acción, concretamente, contra lo que Frankfurt define como el “modelo de deseos y creencias” (1998 y 1998d). Frente

¹³³ Frankfurt no utiliza la expresión “estratégica-instrumental”, pero creo que en este contexto me es lícito hacer uso de ella, teniendo como trasfondo la exposición que se ha hecho sobre Gauthier.

a dicho modelo, creo que resulta de especial interés la ya mencionada distinción que el autor establece entre deseos de primer y de segundo orden, en tanto que ésta admitiría ser vista como una suerte de clave para una “apertura”; para un tipo de “desvelamiento” que arrojaría luz sobre unos terrenos que cabría explorar y analizar, o cuyos elementos y minucias podrían dejar de estar ocultos bajo la escueta y acaso opaca fórmula “A desea X”. Pues ésta podría prestarse para que quede sin ser tematizada, para que permanezca cubierta o pase inadvertida una amplísima gama de posibilidades atribuibles al “desear” humano, y a los posibles y seguramente muy diversos objetos del deseo humano.

Esta amplísima gama de posibilidades tal vez quedaría sin ser considerada si se asumiera demasiado en serio la opacidad de la fórmula misma que, según la crítica de Frankfurt, podría propiciar que tanto el agente A, como el deseo X aparezcan, en principio, como *data* inanalizables, dentro del marco proporcionado por la ya clásica formulación de la teoría contemporánea de la acción: un agente A lleva a cabo una acción B porque A tiene un deseo X, y tiene la creencia Y de que B será un buen medio de obtener la satisfacción de X.¹³⁴ Para Frankfurt, si fueran atendidas las diversas posibilidades que caben para pensar todo aquello que podría ser incluido dentro de la clase de los deseos atribuibles a un agente, se caería en cuenta de la compleja estructura volitiva que el autor le atribuye a aquellos seres a los que llamamos “personas”. Podría, pues, advertirse que tanto los agentes-deseantes, como los deseos que son nombrados en la escueta proposición “A desea X”, son entidades mucho más complejas de lo que la mera fórmula permite percibir a primera vista. Así, X puede designar objetos de deseo de naturaleza muy diversa. Atiéndase, sólo a modo de ejemplo, a las diferencias que habría entre los contenidos de los deseos que se mencionan en las siguientes proposiciones: “María desea *que Pedro la ame*”; “Deseo *beber agua muy fría*”; “El público desea *que no llueva* esta tarde durante el partido”; “María desea *no amar a Pedro*, pues cree que eso le va a hacer mucho daño”; “Pedro desea *que le guste su nuevo jefe, aunque no le gusta nada*, ya que no tiene otro remedio que soportárselo”.

Estas frases sólo son una ínfima muestra de lo que podría ser un infinito o por lo menos bastante numeroso listado de deseos de las clases más variopintas, y entre los

¹³⁴ Cf. Davidson (2001). Remito al lector a la paráfrasis que, en el primer capítulo, hago de esta misma formulación. En “Freedom of the will...”, Frankfurt no hace una alusión directa ni a Davidson ni a otros autores que han contribuido a la teoría contemporánea de la acción, pero sí lo hace en otros textos suyos que en breve comentaré.

cuales se pueden, a su vez, tejer los más diversos tipos de vínculos, al tiempo que cada deseo admite innumerables formas de relación con los demás contenidos intencionales de cada agente. El caso es que los muchos deseos humanos podrían tener naturalezas tan variadas, tan irreducibles las unas a las otras que, por tanto, en principio no tendrían que ser englobados o resumidos en una sola, lo cual, a su vez, es un indicador de la complejidad volitiva de los agentes-que-pueden-desear.¹³⁵ Ahora bien, si intentando reducir un poco dicha complejidad, elegimos entre todas las clases posibles de deseos únicamente aquellos cuyo objeto sean *acciones*, esto es, eso que acaso pueda designar el aspecto más “visible” de los deseos de un agente A, de todas maneras sigue ocurriendo que la frase “A *desea* X” no nos dice mucho acerca de las posibles maneras en que A desea X. Tampoco nos muestra gran cosa sobre las propias *acciones* que efectivamente lleva a cabo A, y menos aún sobre las acciones que A *trata* de realizar. Nos dice aún menos acerca de las *relaciones* que podrían haber entre lo que *de hecho hace* A, lo que A *quisiera hacer* y lo que *intenta seriamente* poner por obra. Tampoco indica nada sobre la *relación* que habría entre las *acciones* que de hecho efectúa A y el *hecho de que* A *desea* X.

Para apreciar un poco más de cerca lo que intenta mostrar Frankfurt, repárese en que el *hecho* mismo de que “A desea X” es un suceso que resulta empíricamente compatible con la ocurrencia de otra gran variedad de *hechos* aparentemente contradictorios con el evento de “desear X”, pero que son posibles como tales hechos en el mundo, a pesar de que es el caso que “A desea X”. Esto puede advertirse, por ejemplo, en los eventos descritos en las siguientes proposiciones: “A desea no desear X”; “A hace no-X”; “A no se ha dado cuenta de que realmente desea X”; “A cree desear X pero también desea no-X”. Como puede verse, X admite ser visto como uno de los muchos deseos que tiene A y, como tal, haría parte de una *estructura compleja* en la que X se interrelaciona con la *evaluación* que A pueda hacer de X, y en la que dicha evaluación también puede presentar diversas facetas e involucrar elementos de naturalezas claramente diferenciables. Esta compleja estructura, pues, puede verse poco o nada reflejada en lo que termina por *hacer* A, esto es, en aquello que, repito, se

¹³⁵ Los ejemplos no son los mismos de Frankfurt, pero creo que con ellos no violento el sentido de lo que en esta parte de su texto él intenta decir.

supone como el aspecto “visible” de su desear: las acciones de A que, también se supone, se explican –al menos en parte– por los deseos de A.

1.1 Deseos “efectivos”, voliciones de segundo orden y el caso de los *wantons*

Un agente humano tiene, entonces, una estructura volitiva mucho más compleja de la que permite visualizar la simple fórmula “A desea X”. De allí que Frankfurt proponga que identifiquemos la “voluntad” o el “querer” (*will*) de un agente humano, con aquello que constituye su “deseo efectivo” (*effective*). Este deseo efectivo, como vemos al examinar someramente parte de todo lo que podría quedar oculto bajo la escueta fórmula “A desea X”, no concuerda necesariamente con muchos de los deseos de primer orden de A, ni tampoco con aquello que meramente “inclina” a A a hacer algo y, ni siquiera, con lo que A “intenta” hacer. Al margen de que A desee X, y de que X sea uno entre sus muchos deseos, si X designa un deseo de primer orden y ocurre que “A desea desear X”, entonces, según Frankfurt, tendríamos que decir que A desea que ese deseo X sea “su voluntad” (*will*). Esto es, A desea que X sea el deseo que efectivamente lo mueva a actuar, A quiere que ese deseo sea su deseo “efectivo” (*effective*), o sea el motor efectivo de su acción, por encima de los demás deseos que tenga A, incluyendo, eventualmente, el deseo (de primer orden) de no-X.

La noción de “deseo efectivo” nos conduce a un elemento adicional de la complejidad de la estructura volitiva de las personas: éstas no sólo pueden tener deseos de segundo orden, como ya hemos visto, sino que, además, pueden tener lo que Frankfurt llama “voliciones” de segundo orden. Alguien que tiene un deseo de segundo orden puede, o bien querer que ese deseo sea su voluntad (*to be his will*), o bien ser indiferente al hecho de que lo sea (o que no lo sea). En el primer caso, advierte Frankfurt, ya no se trata únicamente de un deseo de segundo orden, sino también de una *volición* de segundo orden, puesto que expresa la preocupación del agente por la “deseabilidad” de sus deseos.¹³⁶ Es aquí, en la posibilidad de tener este tipo de *voliciones*, y no sólo deseos de segundo orden, donde Frankfurt sitúa la característica fundamental de la estructura de la voluntad de aquel tipo de agentes humanos a los que estamos dispuestos, según él, a llamar “personas”. Un niño muy pequeño puede tener

¹³⁶ En esta parte sigo “Freedom of the will...”.

deseos de segundo orden, pero éstos no se constituyen propiamente en voliciones de segundo orden. Puede, por ejemplo, no desear el desear el juguete de su vecino, porque puede ganarse una buena reprimenda. Por contraste, un adulto podría desear el desear ser amable y paciente con su vecino, que le parece poco simpático, porque quisiera ser genuinamente movido por esas actitudes y porque quisiera suavizar las asperezas de su carácter.¹³⁷

En el primer caso nos hallamos ante un agente que meramente puede sentir la necesidad de no identificarse con un deseo de primer orden y, por ello, puede tener un deseo de segundo orden sobre su deseo inicial. Mientras que en el segundo caso se trataría de un agente que está preocupado por llegar a ser genuinamente movido por un deseo con el cual, según Frankfurt, quiere *identificarse*, o al que juzga como *más deseable*. Esto no significa que todos los adultos sean propiamente “personas” en el sentido de que necesariamente y en todo momento vivan preocupados por la deseabilidad de sus motivaciones y deseos. De hecho, nuestro autor advierte que los adultos a veces nos comportamos como tales personas, pero otras veces no lo hacemos, ya que en muchas ocasiones no estamos preocupados por la calidad de nuestras motivaciones o por la deseabilidad de aquello que nos mueve a actuar. Sin embargo, los niños muy pequeños sí suelen ser lo que nuestro autor llama *wantons*, es decir, agentes despreocupados por la calidad de sus deseos.¹³⁸

Frankfurt ilustra esta diferencia de radicalidad o de profundidad entre los deseos y las voliciones de segundo orden, comparando lo que sucedería en el caso de dos tipos de drogadictos, uno que lucha contra su adicción, aunque muchas veces sucumba ante ella, y otro adicto al que le tiene sin cuidado cuál debería ser el motivo que le impulse a actuar. Este último tipo de adicto bien podría experimentar, como el primero, un

¹³⁷ Los ejemplos son míos, no de Frankfurt, pero creo que él estaría de acuerdo, por más que, como veremos más adelante, prefiere evitar alusiones a temas relacionados con lo moral o con lo ético. De nuevo, remito al lector a la advertencia que, en un pie de página de la introducción de la tesis, hago sobre los usos de los términos “moral” y “ética”.

¹³⁸ He decidido traducir *wanton* como “despreocupado”, si bien esta expresión está lejos de ser una traducción literal. Espero que el lector comprenda que he privilegiado el sentido de lo que creo intenta decir Frankfurt y también significa el propio término *wanton*, intraducible, en mi opinión, a una palabra castellana que realmente dé cuenta del uso que suele hacerse de él en el idioma original. Este uso refiere tanto a la falta de autorrestricción y de autoevaluación, como a la falta de estabilidad *i.e* actuar según lo dicten los deseos de cada momento y no según una cierta “política” a largo plazo y que indique firmeza o, repito, estabilidad. En lo que sigue, tal vez al lector se le hará más fácil comprender esta licencia que me he permitido con la traducción.

conflicto entre dos de sus deseos: un deseo –de primer orden– de consumir la droga, y un deseo –de segundo orden– de *no desear* consumirla. Sin embargo, afirma Frankfurt, la falta de preocupación por la *deseabilidad* de sus deseos hace que, tratándose del segundo tipo de adicto, el mencionado conflicto entre sus deseos sólo tenga lugar si se presentaran ciertas circunstancias que le impidiesen satisfacer su deseo de consumir la droga y se viera, por consiguiente, en la necesidad de posponer tal satisfacción. Es posible, entonces, que también en su caso hubiera un conflicto entre sus deseos y que el adicto “despreocupado” o *wanton* también tuviese un deseo de segundo orden de no desear la droga, pero dicho metadeseo, según Frankfurt, sólo sería accidental, ya que únicamente se debería a las circunstancias externas. No obedecería en absoluto, insiste nuestro autor, a que el adicto despreocupado se encuentre *concernido* por la *deseabilidad* de sus deseos como, en cambio, sí le ocurre al adicto contrariado. Este último experimenta una lucha entre ambos tipos de deseos, pero esta lucha es de una naturaleza distinta, no es accidental pues, a diferencia del adicto *wanton*, el preocupado se siente *concernido* por cuál *deba* ser el motivo que domine su voluntad.

Por contra de lo que sucede con el adicto despreocupado, el adicto que quiere luchar contra su adicción se encuentra involucrado en una batalla ante la cual no es neutral, puesto que genuinamente quisiera que fuera sólo uno de esos dos deseos –el de no consumir la droga– el que tomara las riendas de su voluntad o fuera su deseo “efectivo”; quisiera, por ende, verse libre de la opresión del deseo contrario. De manera que, tratándose de este tipo de adicto, y a diferencia del adicto despreocupado, puede hablarse de un deseo al que se experimenta como “opresor” y, en su lucha contra él, aunque muchas veces el adicto contrariado “pierda”, por contra, el adicto despreocupado ni ganará ni perderá, puesto que no se halla realmente involucrado en un conflicto, al no estar *concernido* por la *deseabilidad* de aquello por lo que quisiera moverse a actuar. En el caso del despreocupado podría, entonces, hablarse de *deseos* de segundo orden, pero no propiamente de *voliciones* de segundo orden, como ocurre con el adicto contrariado que lucha contra su adicción.

Ahora bien, es importante resaltar que, para Frankfurt, la ausencia de voliciones de segundo orden no implica que el agente tipo *wanton* sea *irracional* o carezca de capacidades deliberativas. Como arriba se menciona, en aquellas ocasiones en las que este tipo de agente experimente un conflicto entre sus deseos, el individuo bien puede

deliberar y decidir posponer la satisfacción de uno de esos deseos encontrados. Lo que hace de este agente propiamente un *wanton* no es la carencia de habilidades racionales, sino el hecho de que no está *concernido* por el tema de la *deseabilidad* de sus deseos. Le tiene sin cuidado la calidad de su voluntad o el asunto de cuál debería ser el tipo de deseo con el que quisiera genuinamente identificarse. El hecho de que opte por un curso de acción antes que por otro, puede que sí involucre una deliberación y una actitud estratégica, pero no una deliberación sobre el tema de la *identidad* que quisiera tener (1998, p. 18). De allí que, según Frankfurt, este tipo de agente no tendría, entonces, más “identidad” que aquella que vaya resultando de la sumatoria de las fuerzas ejercidas por sus deseos de primer orden.

¿Qué papel cumplen entonces sus habilidades deliberativas de las cuales es una prueba el hecho de que tiene deseos de segundo orden? Según Frankfurt, en el caso de un *wanton*, estos últimos únicamente cumplen una función estratégica de orden, de economía o de gestión racional de sus –repito, también eventualmente contradictorios– deseos de primer orden. Empero, sus deseos de segundo orden no son la expresión de la voluntad con la que desearía identificarse, aquella que quisiera poner por encima de sus demás deseos. En este punto es de esperar que el lector se pregunte qué concepto de “identidad” es aquel que Frankfurt parece tener en mente cuando hace estas afirmaciones. De manera que, a continuación me detendré un poco en dicho concepto que, como se irá viendo en lo que sigue, constituye una pieza de gran importancia dentro del esquema propuesto por nuestro autor, si bien no resulta nada sencillo aclarar la naturaleza de aquello que Frankfurt entiende por “identidad”, y por causa de esta idea el autor haya recibido críticas difíciles de sortear.

1.2 Preocupaciones, identificación, “unidad” de agencia y libertad de la voluntad

Para Frankfurt, una vida humana admite ser vista como una *unidad* que va más allá de un mero transcurrir a lo largo del tiempo y de allí que una persona sea mucho más que el *locus* en el que se van sucediendo deseos y otro tipo de eventos mentales (1998a, 1998b y 1998e).¹³⁹ Por esta razón, Frankfurt cree que es posible hablar de un

¹³⁹ Sería interesante intentar establecer una comparación con la idea de “persona” que propone D. Parfit (1984), así como con su tesis de que la permanencia en el tiempo sería más importante que la

agente “integrado”, que contrastaría con una suerte de *wanton*, cuya posible unidad de agencia se desharía en medio de un torrente de deseos que se suceden sin que entre ellos medie una preocupación, por parte del agente, acerca de cuáles de éstos se adecuan a aquello con lo que éste quiere *identificarse*, *i.e* quiere asumir como motor real de sus acciones.¹⁴⁰ Ahora bien, lo que permitiría, según Frankfurt, que una vida humana sea un todo o una unidad “integrada” es aquello a lo que nuestro autor se refiere como las *preocupaciones* fundamentales (*caring*) de una persona. A lo largo de su vida, el individuo experimenta y manifiesta unas ciertas preocupaciones que constituyen el núcleo de temas a partir de los cuales la persona “organiza” su vida, hace planes, emprende acciones, se traza metas y establece prioridades. Lo cual, por supuesto, estaría estrechamente relacionado con la evaluación de sus deseos y la toma de sus decisiones, así como con el importante hecho de que, por la vía de dicha evaluación, el agente se “identifica” con algunos de sus deseos y, por el contrario, rechaza otros, tal y como ya se mencionó.

De este modo, según Frankfurt, las preocupaciones de una persona permiten que su vida sea propiamente “suya” y no se reduzca a un transcurrir de eventos que

identificación con ciertos contenidos mentales, entre los cuales, por supuesto, entrarían los deseos a los que Frankfurt se refiere. Pero creo que ello supondría traspasar los límites de esta exposición. Con relación a este tema también son pertinentes las objeciones planteadas por J. D. Velleman (2002) en contra de lo que éste ve como una exagerada y poco sana –incluso neurótica– preocupación de Frankfurt por la “unidad” de la persona. Sobre esta última crítica volveremos reiteradamente.

¹⁴⁰ Alguien podría objetarle a Frankfurt que no sólo la persona preocupada por el tema de su identidad y por la deseabilidad de sus deseos presenta esta “unidad” a la que este autor se refiere. De hecho es perfectamente posible observar una clara unidad o una coherencia en la conducta de un individuo que está obsesionado por algo ante lo cual no asume en absoluto la distancia crítica y/o reflexiva implicada en el tipo de evaluación que, según Frankfurt, caracteriza a una “persona” y la distingue de un *wanton*. El caso del adicto despreocupado podría ser un ejemplo de ello. Incluso, podría irse más allá y señalarse que la conducta de algunos agentes que presentan patologías mentales suele ser bastante consistente, “unificada” y previsible. Ante este tipo de objeciones, creo que el autor nos remitiría a algo que ya se ha mencionado arriba sobre la ausencia de auténticas voliciones de segundo orden en los *wantons*. Así, Frankfurt tal vez nos diría que esta forma de conducta propia de agentes obsesionados y/o alienados podría dar la impresión externa de coherencia, pero que dicha coherencia sería, según nuestro filósofo, únicamente accidental. Sería el resultado fortuito de un modo de agencia que no cuenta con una unidad asumida y querida “desde dentro” por el propio agente que se preocupe por intentar obtener dicha unidad, y de allí que pueda observarse “desde fuera” como tal unidad. Es posible que esta respuesta no sea la que ofrecería Frankfurt, pero me aventuro a decir que sería consistente con lo que se alcanza a ver en su discurso. También supongo que no es una respuesta satisfactoria, lo cual se puede apreciar mejor a la luz de críticas como las que le hace Velleman (2002) a nuestro autor. Igualmente, este asunto está relacionado con algunas dicotomías que parece asumir Frankfurt, *v.g.* entre “lo externo” y lo propiamente “interno” a la voluntad del sujeto, la “pasividad” de éste *vs.* su “actividad”, asunto que también puede ser bastante problemático y, en mi opinión, no queda claramente zanjado en las respuestas de Frankfurt a sus críticos (2002d y 2002f), tal como puede verse en el caso de las objeciones de R. Moran (2002) y de T. M. Scanlon (2002).

meramente “le suceden” al agente y éste se limite simplemente a sufrir pasivamente. Lo cual, para nuestro autor, nos hace ver que algunos de los deseos experimentados por el agente son susceptibles de ser tomados por éste como propiamente “suyos”, esto es, como algo por lo cual el agente asume una responsabilidad, al integrarlos como parte de la vida que el agente quiere vivir. Esta responsabilidad se relaciona con otra importante distinción establecida por Frankfurt: existe, según él, una diferencia entre la mera libertad de acción y aquello que nuestro autor llama “libertad de la voluntad”. Cualquier agente, incluso uno que no pertenezca a la especie humana, puede gozar del primer tipo de libertad y, en virtud de ésta, podría actuar y moverse según lo desee, siempre y cuando no se le presente el obstáculo impuesto por alguna interferencia externa al propio agente.

En contraste con este tipo de libertad, nuestro autor cree que la libertad de la *voluntad* requiere que el individuo logre actuar según el tipo de deseos con los realmente *se identifica*,¹⁴¹ en el sentido de que sus deseos efectivos sean precisamente aquellos que se adecuen a sus voliciones de segundo orden, según el esquema arriba comentado. Habría pues, una estrecha conexión entre este tipo de libertad de la que goza un agente cuando *hace* lo que *quiere desear y hacer*, y el núcleo de sus *preocupaciones* fundamentales. Éstas explicarían por qué el agente se “identifica” con algunos de sus deseos y, en cambio, rechaza aquellos otros deseos que contradigan dichas preocupaciones. Un agente que logra actuar en razón de tales preocupaciones y, en virtud de éstas, consigue que sus deseos efectivos obedezcan a sus voliciones de segundo orden, conserva, según Frankfurt, una “unidad” de agencia a lo largo de su vida. De esta manera, la persona logra que su compromiso con una identidad no se deshaga a pesar de los cambios que se operen en las circunstancias por las que vaya atravesando su vida. Por lo anterior, el núcleo de las preocupaciones importantes de un individuo es aquello que permite que su vida se “unifique” o se “integre (1998b).

Este asunto de las preocupaciones es introducido por Frankfurt como una novedad en el contexto de los temas recurrentes que nuestro autor le atribuye a la tradición filosófica o a los trabajos comúnmente desarrollados por los filósofos. En

¹⁴¹ En este punto seguramente el lector podrá apreciar cierta relación con lo que B. Williams (1995) llama “razones internas”. De manera parecida a lo que sucede con Parfit, creo que sería interesante ahondar un poco más en las posibles similitudes con Williams, pero pienso que esto no es del todo necesario y traspasaría los límites dentro en los que por ahora quisiera enmarcar este análisis: la propuesta de Frankfurt.

artículos como “The Importance...” (1998d), y “On Caring” (2007c),¹⁴² Frankfurt afirma que la filosofía se ha venido ocupando únicamente de dos tipos básicos de problemas, a saber, los de la epistemología, que giran en torno de la pregunta “¿qué creer?”, y los de la ética, que se originan en la cuestión “¿cómo comportarnos?”. Este monopolio compartido por la ética y la epistemología ha originado, en opinión de Frankfurt, el inconveniente de que los filósofos han terminado por descuidar e incluso por ni siquiera caer en cuenta de que hay un tercer grupo de problemas que se originarían en el intento por responder a la pregunta “¿sobre qué preocuparnos?” (*what to care about?*). Buena parte de la argumentación de Frankfurt estará, entonces, orientada a mostrar la plausibilidad de un discurso filosófico, novedoso según él, que se encargue de lo relacionado con esta tercera pregunta. Este discurso, afirma nuestro autor, se hace necesario tanto por la relevancia misma de su asunto, como por el hecho que éste no podría reducirse ni a las cuestiones epistemológicas, ni a las cuestiones éticas –también llamadas por Frankfurt “morales”–, puesto que, en opinión de nuestro filósofo, desde ninguna de esas dos ramas de la filosofía tradicional puede darse cuenta del hecho de que las personas somos seres para los cuales las cosas importan: “Those disciplines need no reflect upon the nature of caring as such, nor they are obliged to consider what is implied by the fact that we are creatures to whom things matter” (1998d, p. 80).

Creo que un lector desprevenido podría sorprenderse ante este juicio negativo de Frankfurt sobre los olvidos que nuestro autor le atribuye a los filósofos y, sobre todo, ante su idea de una necesaria escisión entre el preocuparse y los asuntos epistemológicos, de un lado, y el preocuparse y los problemas morales, del otro. Pues pienso que, si bien podría admitirse que la epistemología no tendría que ser considerada como la fuente de todas las respuestas a la pregunta por eso que nuestro autor parece entender por “preocupaciones humanas”, por lo menos sí contendría muchos temas que se solaparían con éstas, v.g. aquellos que tienen que ver con valores tales como el rigor

¹⁴² “On Caring” fue traducido como “Sobre el preocuparse” y hace parte de una compilación que en Castellano fue publicada en 2007 bajo el título de “Necesidad, volición y amor”. La versión original de dicha compilación data del año 2006. En ella se incluyen textos de los años 60s, 70s, 80s, y 90s, pero los artículos a los que haré referencia son fundamentalmente de los 80s y 90s, siendo aquellos en los que nuestro autor vuelve sobre los temas que son objeto de este capítulo, temas sobre los ya había disertado en los 70s y con relación a los cuales creo que uno podría establecer una comparación en cuanto a la persistencia de ciertas preocupaciones, y en cuanto a las continuidades o cambios que puedan advertirse en las tesis sostenidas por Frankfurt.

en la investigación, o con la preocupación por la verdad.¹⁴³ Así mismo, pienso que también sorprende que nuestro autor excluya a la ética (repito, a la que Frankfurt, por lo menos en muchos de sus textos, no distingue de la moral) del tipo de asuntos por los cuales una persona pueda preocuparse. Frankfurt insiste en que la ética/moral sólo tiene que ver con aquello que resulta ser correcto o incorrecto, justo o injusto, en nuestras relaciones con *las demás personas*, no con lo que quisiéramos hacer con *nosotros mismos*, no con aquello a lo que quisiéramos otorgarle importancia en nuestras vidas. Pero ocurre, según nuestro autor, que la gran importancia que los filósofos le otorgan a la moral no hace justicia a un hecho innegable: somos seres que estamos preocupados, a los que importan una gran cantidad de asuntos que no se reducen en absoluto a los problemas éticos. Para Frankfurt, nuestras vidas no se realizan única ni preferentemente en orden a satisfacer las demandas de la ética, esto es, lo que, repito, nuestro autor entiende como aquellas normas que regulan las relaciones de justicia o corrección moral con las otras personas. De nuevo, dejo por ahora simplemente anotada la perplejidad que podría asaltar a algún lector que se encuentre con estas separaciones tan radicales que Frankfurt establece entre las preocupaciones de las personas y los asuntos que puedan revestir alguna relevancia que el autor calificaría de “epistemológica”, y entre las mencionadas preocupaciones y aquello que quepa considerar como de importancia moral, sin mencionar su definición de “lo moral” y el hecho de que lo haga indistinguible de “lo ético”. Este tema será retomado, a la vista de la importancia que reviste y que, como veremos, hará necesario que se evalúen sus consecuencias para la plausibilidad misma de la propuesta de Frankfurt. Permítaseme, por lo pronto, continuar con una de las implicaciones más importantes que tendría para nuestro autor el papel central que él le atribuye a las preocupaciones de una persona: aquello que Frankfurt llama “necesidad volitiva”.

¹⁴³ Al respecto, véase: Williams (2002). En la noción de “valor” que parece manejar Frankfurt, también me detendré más adelante.

2. La “necesidad volitiva” y lo “impensable”

Uno de los rasgos más interesantes de las preocupaciones de una persona es que, a pesar de que actuar por mor de ellas es la base de su autonomía *i.e.* de que logre moverse a actuar únicamente por aquellos deseos que quiere que sean sus deseos efectivos, aquellos con los que realmente se identifica, sin embargo, según Frankfurt, el *caring* de un agente no se encuentra del todo bajo su control. La persona no puede preocuparse por algo simplemente porque de manera voluntaria así lo decida. Incluso pareciera que de algún modo sus preocupaciones “se le imponen”. De allí que nuestro autor crea que es posible hablar de un cierto tipo especial de “necesidad” que no es una necesidad lógica, ni tampoco causal, sino volitiva, y que se advierte, precisamente, en ese tipo singular de temas que preocupan a las personas, temas a los que, como veremos, Frankfurt se resiste a relacionar con una cuestión de “valor”.¹⁴⁴ Aquí, de nuevo, creo que puede causar extrañeza un hecho que a primera vista podría ser considerado como contradictorio: que nuestro autor, de un lado, sostenga que las preocupaciones de una persona no son algo que ésta elija voluntariamente, y de otra parte, que al mismo tiempo Frankfurt insista en no involucrar dichas preocupaciones con una cuestión de “valor”.

Hablo de una posible contradicción, pues la primera afirmación bien podría indicar que el agente asocia las preocupaciones con una suerte de referente *normativo* que, como tal, tendría el tipo de estatus de intersubjetividad que normalmente se atribuye a las cuestiones que revisten importancia para una persona independientemente

¹⁴⁴ “Thus caring about something is not to be confused with liking it or with wanting it; nor is it the same as *thinking* that what is cared about has *value* of some kind, or that it is *desiderable*”. (1998d, p. 83; las cursivas son mías). Como vemos en esta afirmación de Frankfurt, y en breve volverá a comentarse, aquí se podría estar haciendo un uso de términos como “deseable” y “pensar”, los cuales me permito resaltar (junto con la palabra “valor”), ya que pienso que dicho uso bien podría prestarse para que el lector se pregunte si no habría una suerte de inconsistencia entre: 1) La tesis anticognitivista que Frankfurt enfáticamente sostiene en este aparte, al referirse a la naturaleza que, según él, debe atribuírsele al *caring*, y 2) La importancia que nuestro autor cree que debe concedérsele a la “evaluación” de los deseos, importancia que, como intentaré mostrar, debería llevarle a suscribir una posición eminentemente *cognitivista* en cuanto a cómo han de concebirse las preocupaciones de las personas y, por ende, en cuanto el “juicio” que creo que tendría que hacer un agente que “evalúe” la “deseabilidad” de sus deseos. Hablo de “juicio”, aunque Frankfurt seguramente no estaría de acuerdo, ya que no veo cómo puede alguien pronunciarse sobre dicha deseabilidad, o sobre sus objetos de preocupación si no los “juzga”, y los juzga precisamente apelando a referentes normativos, con relación a los cuales dichos objetos son vistos como poseyendo un “valor” que explicaría el hecho de que la persona se preocupe por ellos, independientemente de que, como dice Frankfurt, “le gusten” o no.

de sus gustos y conveniencias,¹⁴⁵ y en las que el agente reconoce una suerte de *autoridad* que no puede menos que ser relacionada con el “valor” que el sujeto ve implicado en dichas cuestiones; sin que esto signifique asumir, como hace Frankfurt, la idea de que el término “valor” implica darle juego a un objetivismo de los valores, estilo Moore o Scheller, tal y como vimos que sucede en el caso de Gauthier. De allí que creo que, como ocurre con el filósofo canadiense, es posible encontrar problemas en el concepto mismo de “valor” que Frankfurt parece asumir como canónico, de modo similar a lo que arriba se señala con respecto a su uso de la noción de “moral”. Como este tema, repito, será objeto de un ulterior desarrollo, en vista de las importantes implicaciones que tiene, no me detengo por ahora más en él, pero sí quería dejar constancia de la extrañeza que puede suscitar.

2.1 El *caring*: libertad y necesidad

Continuando con el asunto de la naturaleza que nuestro autor atribuye a las preocupaciones, Frankfurt considera que éstas de algún modo determinan una clase de asuntos que *se nos imponen* como exigencias que no nos sentimos capaces de contravenir y, lo que es más interesante, contra las cuales tampoco queremos resistirnos. Nos sometemos voluntariamente a ellas, de manera análoga a como nos sentimos subyugados por la belleza (1998d y 2007c). De allí que nuestro autor vea en ello la evidencia de una necesidad a la que, repito, él califica de “volitiva”. Tal vez el ejemplo más diciente con el que Frankfurt trata de ilustrar esta necesidad, la cual estaría presente en las preocupaciones fundamentales de las personas, sea el de la actitud de Lutero, reflejada en su famosa declaración “Aquí estoy, no puedo hacer otra cosa”.¹⁴⁶ Este “no poder” implica que hacer lo contrario de aquello que el agente se siente obligado a hacer, en virtud del tipo especial de necesidad que proviene de sus preocupaciones, de aquello que considera de fundamental importancia para su vida, significaría para esta

¹⁴⁵ Tal vez en este punto se podría establecer una relación entre aquello que Tugendhat (1988) vería como una cierta “pretensión de objetividad”, y Habermas (1989 y 1994a) considera una “pretensión de validez” por parte del agente que emite un juicio relacionado con lo normativo-moral. Obviamente ello no significa que *ipso facto* el juicio cuente con la aceptación cualificada de los demás agentes, pero sí implica que el sujeto en cuestión cree que dicho juicio se encuentra respaldado por buenas razones, razones que no tienen que ver con sus gustos e intereses meramente personales y, por ende, que el agente honestamente asume que su juicio cuenta con los requisitos para obtener tal aceptación intersubjetiva.

¹⁴⁶ Citado por Frankfurt en: 1998d, p. 86.

persona hacer lo “impensable” (*unthinkable*). Esto es, lo que se sale por completo del marco de las acciones que el agente estaría dispuesto a llevar a cabo, e incluso, indicaría el tipo de actos que la persona no se siente capaz de *concebir* o de tan siquiera tomar en consideración (1998c).¹⁴⁷

La “necesidad” que parece imponerse a la voluntad de una persona que no quiere traicionar su *caring*, es cierto que no resulta ser la misma que aquella que caracteriza otros órdenes dentro de los cuales también cabe hablar de “necesidad”, tales como las relaciones causales y las conceptuales. Empero, según Frankfurt, sí se trataría de una necesidad que resulta de algún modo análoga a las necesidades que se advierten en dichos órdenes. Pues en el caso de las voliciones humanas también habría una suerte de “imposibilidad” de traspasar ciertos límites que el sujeto no sólo no quiere irrespetar, sino que, además, se trata de límites cuya trasgresión es percibida por el agente como impensable, esto es, como una suerte de “imposible” que se asemeja más a la imposibilidad lógico-conceptual, que a un imposible empírico. De allí que Frankfurt también intente explicar la naturaleza de esta necesidad volitiva comparándola con aquella necesidad lógica o conceptual que explicaría, por ejemplo, que el concepto mismo de “triángulo” implique necesariamente la imposibilidad de que la suma de los ángulos interiores de un triángulo no sea de 180 grados (2007 y 2007a).

De un modo similar a como no nos es posible pensar que tal suma sea distinta a 180 grados, una persona tampoco puede sentirse capaz de considerar llevar a cabo cierto tipo de acciones, las cuales sencillamente no puede decidirse a realizar e, incluso, ocurre que ni siquiera se puede ver a sí misma como llevándolas a cabo. Se trata de acciones que el sujeto percibe como inconcebibles para sí mismo, como trasgresoras de unos límites que no son otros que los de su propia “unidad” como persona. Según Frankfurt, si estos límites se quebrantan, habría una traición en contra de lo que el sujeto concibe como el centro de sí mismo, esto es, lo que en sus textos más recientes, nuestro autor llama la “esencia” o la “naturaleza esencial” de la persona (2007 y 2007a).¹⁴⁸ De allí que dicha traición, a los ojos de Frankfurt, no deba ser considerada como equiparable al

¹⁴⁷ Al respecto, véase el comentario de Scanlon (2002) sobre esta imposibilidad de “considerar”.

¹⁴⁸ Sobre el problema de pensar que tenemos una “esencia” y, por consiguiente, creer que, si ésta cambia, entonces “nos rompemos” como las personas que somos, resulta especialmente ilustrativa la crítica de Velleman (2002) que, como advertí en un pie de página anterior, se mencionará reiteradamente a lo largo de este capítulo.

tipo de ruptura que, en otros contextos y tratándose de otro tipo de asuntos, podría ser vista como un acto emancipatorio o liberador. Antes, por el contrario, según nuestro autor, hacer lo impensable consistiría en un ataque en contra de la propia libertad de la voluntad del agente.

De manera que el concepto mismo de “libertad” parece estar implicado, paradójicamente, en esta noción propuesta por Frankfurt, de una “necesidad volitiva”. Esto acaso se entienda mejor si se lo expresa en términos positivos. Volvamos al caso de Lutero: si este personaje dice que “no puede” hacer otra cosa, no es porque empíricamente no tiene más opciones, ni tampoco, y lo que es más importante, porque se sienta oprimido por una fuerza externa que le obliga a hacer lo que él realmente no quiere. Por el contrario, Frankfurt cree que la suerte de “fuerza” a la que se alude en este caso, proviene de aquello con lo cual el agente se identifica profundamente como algo definitorio de su propio yo (*self*). Esto es, se trata de una suerte de fuerza que se debe a su propio *caring*, y en éste, recordemos, se sintetiza lo que la persona más quiere hacer y más quiere desear. Ahora bien, como ya se ha dicho, si las preocupaciones fundamentales de una persona no son elegidas a capricho por el sujeto, entonces en el *caring* y en el tipo de “autoridad” (2007a y 2007c)¹⁴⁹ que emana de sus exigencias, confluyen sin violencia dos aspectos aparentemente contradictorios: la libertad y la necesidad. Estaríamos hablando de una libertad en virtud de la cual la persona se libera de fuerzas externas a su voluntad, fuerzas con las que no se identifica y que resiente como opresivas. Esta libertad igualmente presenta el aspecto positivo del actuar por mor de aquellos deseos que la persona quiere que sean el motor de su acción. Y en cuanto a la necesidad que se entrelaza con dicha libertad, se trataría de aquel tipo de necesidad marcada por unos límites que la persona no se siente capaz de traspasar: los límites de su propia voluntad, de lo que la persona se siente capaz de querer e, incluso, de concebir.

Por lo anterior, el ejemplo de Lutero podría ser visto como el caso contrario al del adicto preocupado, esto es, aquel tipo de agente que se siente oprimido por fuerzas externas que se le imponen aunque él no se identifique con ellas, fuerzas que quisiera poder resistir o vencer, y a las que no quiere asumir como su deseo efectivo o su

¹⁴⁹ Frankfurt utiliza la palabra “autoridad” relacionándola, como veremos, con el tema del “amor” que, según él, una persona siente por aquellos seres o aquellos asuntos por los que más se preocupa.

voluntad. Pero en el caso de Lutero, por el contrario, no se trataría en absoluto de fuerzas externas y opresivas, sino de aquello que el agente considera de la mayor importancia y con lo cual se siente profundamente identificado. Por esto se siente igualmente obligado a no hacer lo que no puede y tampoco quiere hacer. Sería, por tanto, el caso de una libertad que, según Frankfurt, estaría imbricada con un sometimiento voluntario a una suerte de “necesidad” en virtud de la cual el agente no puede contravenir eso por lo cual está radicalmente preocupado; aquello con lo que se siente tan identificado que su contravención sería algo verdaderamente inconcebible para el sujeto, o estaría por fuera de toda consideración posible, esto es, sería “impensable” (*unthinkable*) para él. En este punto, es importante señalar de nuevo un asunto que ya se ha mencionado: Frankfurt cree que cuando las personas actúan de acuerdo con esa necesidad volitiva, o buscando no hacer aquello que consideran impensable para ellas, no es porque estén respondiendo a un problema *moral*, sino sólo a sus preocupaciones más importantes, las cuales, recordemos, según nuestro autor son irreducibles a lo moral. Arriba ya se ha comentado la extrañeza que pueda sentir un lector que se encuentre con esta separación tan radical entre el *caring* y lo moral, pero creo que dicha extrañeza se puede agudizar si se repara en los ejemplos a los que acude Frankfurt para ilustrar lo que él entiende como el tipo de necesidad volitiva que nace de las preocupaciones de una persona, y la irreducibilidad de “lo impensable” a una contravención de normas morales. Ya se ha mencionado el caso de Lutero, el cual, en mi opinión resulta difícil no leer en clave moral ni ética.

Pero, además de éste, nuestro autor acude a dos ejemplos que creo producirían el efecto contrario al que deseaba Frankfurt. Pienso que, con los ejemplos elegidos, nuestro autor en vez de lograr que el lector se convenza de que éstos son claras instancias de que las preocupaciones excluyen lo moral/ético, el lector vería en tales ejemplos la prueba de la tesis contraria a la que sostiene Frankfurt. En primer término, el autor propone el ejemplo de una madre que se enfrenta a la posibilidad de entregar a su hijo en adopción y, aun cuando ella vea que esta sea la salida más conveniente para el niño, sin embargo, no se siente capaz de abandonarle: este acto es impensable para ella. La madre sencillamente no puede dejar al hijo, pero, insiste Frankfurt, no es algo que ella rechace porque se sienta atada a algún tipo de obligación moral, sino porque sus más fundamentales preocupaciones le hacen sentir la fuerza de esa necesidad volitiva en

virtud de la cual ella simplemente “no puede” abandonar al hijo, o por la cual llevar a cabo ese acto es visto por ella como algo “impensable” para sí misma (1998d, p. 90).¹⁵⁰ Igual lectura hace Frankfurt del conocido ejemplo propuesto por Sartre en *El existencialismo es un humanismo*: el caso de un muchacho que, durante la invasión alemana a Francia, se debate entre quedarse a cuidar a su madre enferma, o irse a combatir del lado de la Resistencia Francesa. Según Frankfurt, lo que el chico finalmente decida no debe ser visto como el producto de una elección moral, sino como un asunto que atañe sus preocupaciones fundamentales, a partir de las cuales terminará por escoger aquello que “pese más” a nivel de dichas preocupaciones, no aquello que vea como el cumplimiento de alguna obligación moral (1998d, p. 84).¹⁵¹

2.2 La racionalidad prediscursiva, el amor y la exclusión de lo normativo

¿Qué es, entonces, lo que habría que entender como propio de las preocupaciones de una persona, si de éstas queda excluido todo elemento moral y si, como vimos, tampoco hay en ellas involucrados ni un juicio de valor, ni la apelación a ningún tipo de referente normativo? ¿Cómo entender que Frankfurt sea tan enfático al afirmar que el hecho de que una persona se preocupe por algo y lo considere, por tanto, de una gran importancia para sí misma, *no* significa que lo juzgue como especialmente “valioso” o “meritorio”? Pienso que esta posición de Frankfurt en contra de introducir lo normativo¹⁵² en el ámbito de las preocupaciones se hace todavía más difícil de

¹⁵⁰ Al respecto, véase las críticas de Bratman (2002) y de Scanlon (2002) a la interpretación que hace Frankfurt de este ejemplo.

¹⁵¹ Es interesante que Frankfurt acuda a los mismos ejemplos en artículos de la época de “The importance...” (1998d), y en sus artículos más recientes, como “Sobre el preocuparse” (2007c), y que en esta última época insista en hacer *la misma* interpretación que hiciera de tales ejemplos dos décadas antes, salvo, tal vez, por la novedad que podría significar que introduzca el tema del amor. Sobre el ejemplo de Sartre volveremos en el quinto capítulo, cuando veamos el análisis que hace Taylor del significado de los dilemas morales.

¹⁵² En claro contraste con Frankfurt, para C. Korsgaard, los referentes normativos son fundamentales en el proceso de construcción de nuestra identidad práctica y moral. Para Korsgaard, nuestra agencia se estructura no sólo a partir de que como agentes se nos identifique como autores de nuestras acciones, ni aun siendo capaces de armonizar en una estructura jerárquica nuestros deseos, sino que son los principios normativos implicados en nuestras decisiones y con los cuales queremos identificarnos, los que moldean nuestra identidad y nos permiten integrar nuestra agencia en una perspectiva moral desde la cual podemos juzgar nuestras acciones como buenas o malas. Como los principios que guían nuestras acciones obedecen a criterios de racionalidad pueden, por ello, ser juzgados de acuerdo a estándares de eficacia (como imperativos hipotéticos), o de autonomía y universalización (como imperativos categóricos), tanto por nosotros como por otros y de allí derivan la fuerza normativa que tienen sobre nuestras acciones.

comprender si se tiene en cuenta otras dos nociones que introduce como novedosas: la idea de una *racionalidad de la voluntad*, de una racionalidad prediscursiva, y su noción del *amor* como centro neurálgico de las preocupaciones. En relación con la *primera* idea, Frankfurt cree que puede hablarse de una forma “básica” de racionalidad que explicaría por qué nos resultan, según él, tan estrafalarias ciertas preferencias, por ejemplo, aquella que manifiesta un sujeto que haga suya la conocida frase de Hume: “no es contrario a la razón el preferir la destrucción del mundo entero a tener un rasguño en mi dedo.”¹⁵³.

En “Rationality and the unthinkable”, Frankfurt intenta demostrar que sí puede hablarse de preferencias irracionales, como la que podría atribuirse al sujeto que lance una frase como la de Hume. Una persona que prefiera que se destruya el mundo a tener un rasguño, según Frankfurt, no sólo tiene un problema moral, sino que, además, es irracional, no porque carezca del tipo de racionalidad propia de quien posee habilidades deliberativas y estratégicas, sino porque presenta fallos en cuanto a una forma de racionalidad que es propia de quien participa del concepto de *lo impensable*, de quien puede ser conciente de ciertos límites impuestos por aquello que sencillamente no puede “querer”, so pena de ignorar esa necesidad que, según lo acabamos de ver, es de la voluntad pero, al mismo tiempo, involucra lo que la persona está en capacidad de concebir con respecto al límite de sus acciones. Según Frankfurt, quien participa de esta idea de lo impensable, también es competente como agente que posee una forma de racionalidad “más básica” que aquella que asociamos a nuestra capacidad discursiva; esto es, una suerte de racionalidad de la voluntad. Pero, ésta, de nuevo, según nuestro autor, no puede ser equiparada ni a un asunto moral, ni a un problema relacionado con nuestros juicios de valor (1998c, pp. 188-190).

¿Cómo entender la naturaleza de esta “racionalidad”? Creo que Frankfurt no responde a esta pregunta en el artículo que acaba de mencionarse, ni en los textos

Además, para Korsgaard, los principios implicados en nuestras decisiones nos obligan, no por una necesidad lógica o causal, ni tampoco por una necesidad racional sino que, por ser incondicionados, proceden de una necesidad interna (*necessitation*) que es espontánea y que revela y expresa el tipo de persona que somos. Es esta espontaneidad la que inicia el proceso de construcción de nuestra identidad como agentes que intervienen libremente en el mundo, pero la construcción de esa identidad es entendida por Korsgaard como autoconstitución (*self-constitution*): es a través de lo que hacemos, y cómo lo hacemos que nos unificamos e integramos como agentes, nos hacemos a nosotros mismos y nos constituimos como personas. Cf. Korsgaard, C. M. (2009).

¹⁵³ *Tratado de la naturaleza humana*, Libro II, Sección III, § 416.

contemporáneos a éste, amén de que en sus trabajos más recientes tampoco menciona el tema y, cuando habla de “racionalidad” parece resistirse a ver el tema del *caring* como relacionado con cualquier posibilidad de que el agente explicita *razones* para su acción.¹⁵⁴ En este punto, creo que se manifiesta una tensión análoga a lo que anteriormente se señaló con respecto a la posible contradicción que habría entre considerar que el *caring* no obedece a una decisión voluntaria de la persona y, al mismo tiempo, asumir un anticognitivismo en virtud del cual se niegue que en las preocupaciones se manifiesta un elemento normativo, llámese valorativo, llámese moral. Esto tal vez se muestre de forma más clara si hacemos una relectura de los mismos ejemplos a los que acude Frankfurt. Si recordamos lo dicho sobre el tema de la identidad del agente preocupado por la deseabilidad de sus deseos o por la calidad de los deseos que quisiera querer, o que quisiera que lo movieran efectivamente a actuar, sobre todo en casos como el de los dos tipos de adictos, creo que no es fácil hacer compatibles: 1) Esta postura de Frankfurt en contra de lo normativo y lo cognitivo, y 2) Lo que el autor sostiene acerca de la deseabilidad de los deseos y de lo que distingue al adicto preocupado.

En mi opinión, dicha postura anticognitivista no encaja con su pintura de los dos adictos, ni con su idea inicial de que las voliciones de segundo orden implican una capacidad para autoevaluarse y criticar los propios deseos. Igualmente, creo que habría dificultades para entender la importancia que Frankfurt le concede a la libertad de la voluntad, si el agente que goza de tal libertad no involucra lo normativo en la formación de sus preocupaciones que serían, precisamente, el centro de dicha libertad, que es la que disfrutamos cuando hacemos y queremos lo que queremos-querer; cuando nuestros deseos efectivos son precisamente aquellos que sí quisiéramos que nos movieran a la acción. Si nuestro autor ha insistido en que los agentes que se comportan como personas son aquellos que evalúan y critican sus deseos, incluso hasta el punto de estar

¹⁵⁴ Al respecto, resulta bastante ilustrativa la polémica que sostiene Frankfurt con autores que, como Bratman (2002), Scanlon (2002) y Watson (2002), le critican esta insistencia suya en no ver el aspecto *cognitivo* y la cuestión *normativa* que se juega en la evaluación de los deseos. Estos autores, ofreciendo diversas razones, pero relacionadas entre sí, intentan mostrarle a Frankfurt que sería impropio ignorar el hecho de que, si mediante dicha evaluación un agente llega a identificarse con algunos de sus deseos y, en cambio rechaza otros como “indeseables”, es porque en ello necesariamente tendría que estar implicada la apelación a algún tipo de *referente normativo*, y el uso de alguna forma de *deliberación*, o de *dar razones*. Creo que resulta llamativo que Frankfurt, a pesar de las críticas lanzadas por estos autores, en sus réplicas a éstas (2002, 2002c y 2002f) se aferre a su postura anticognitivista e insista en seguir negando el papel que jugarían las “razones” y lo normativo dentro del ámbito de las preocupaciones.

preocupados por la deseabilidad de éstos, o por identificarse con algunos deseos que quisieran que fueran sus motivos para actuar, pienso que Frankfurt debería aclarar, si no se trata de un referente moral, entonces ¿cuál es la naturaleza del referente –pienso que evidentemente normativo– que les hace considerar a las personas que algunos de sus deseos son “deseables” y otros, por el contrario, son “indeseables”? ¿Qué tipo de referente normativo tiene, entonces, en mente, el adicto que lucha contra su adicción? ¿Por mor de qué se involucra en esa lucha? ¿Por qué tipo de razones quiere el adicto contrariado que cambien sus deseos, o que sólo uno de ellos sea el que lo mueva efectivamente a actuar y lo mueva, precisamente, por encima o en contra de otros deseos a los que juzga como indeseables? ¿En razón de qué los considera “indeseables”? Creo que este tipo de preocupación, la que distingue a una persona de un *wanton*, y que está estrechamente vinculada con la alta valoración que, según Frankfurt, le damos y debemos darle los agentes humanos a nuestra libertad de la voluntad (por encima incluso de nuestra libertad de acción), no se entiende fácilmente si no hay en ello involucrada una cuestión normativa que, por ahora, no se ve por qué no pueda incluir preocupaciones morales (expresión acaso anómala a los ojos de Frankfurt) y también aquellas que, según lo visto en la introducción de la tesis, podríamos llamar “éticas”. Esto, sin mencionar que el agente –por qué no–, podría también estar preocupado por valores de carácter epistémico.

Pienso que estas inconsistencias originadas en el anticognitivismo de Frankfurt, terminan por ser un pesado lastre para que su discurso fluya, se desarrolle hasta sus últimas consecuencias, se haga más claro y convincente y, por ende, logre también sortear las objeciones lanzadas por algunos de sus críticos. Adicionalmente, creo que tales inconsistencias y la prohibición de introducir elementos normativos en las preocupaciones de las personas, incluyendo el elemento moral, terminan por truncar uno los avances o de las novedades que cabría esperarse de Frankfurt: que su idea de agencia logre desprenderse efectivamente de modelos como el que vimos ejemplificado en Gauthier. Con ello, se esperaría que, una vez que se han sentado bases tan importantes como el tema de la evaluación de los deseos, igualmente se logre, parafraseando la expresión de Nozick, “ir más allá de Hume”.¹⁵⁵ Esto es, se logre disolver la opacidad de la base motivacional atribuible a los agentes humanos. Sin

¹⁵⁵ Nozick (1995, p. 91).

embargo, creo que este objetivo no se alcanza del todo, y el lector quedaría *ad portas* de verlo logrado. Lo cual se agrava aun más por la introducción de aquello que, en las últimas obras de Frankfurt, es señalado por éste como el corazón de las preocupaciones de la persona: el amor que, según nuestro autor, ésta siente por aquellos individuos y aquellos asuntos por los que más se preocupa (2004, 2007a y 2007c).

Este amor del que habla Frankfurt no se reduce a la pasión relacionada con la atracción sexual o los vínculos sentimentales que unen a una pareja, sino que es un amor que obedece a la voluntad. De manera que no se relaciona con los elementos “pasivos” de nuestra psique, tales como las pasiones que se salen de nuestro control. De allí que también sea un amor extensible a otros objetos, además de las personas que amamos, por ejemplo, a una causa, a nuestra patria, etc. Se trata de un sentimiento altruista, en el que lo importante no es el beneficio ni los placeres que obtenga el amante de su amado, sino que, por el contrario, el bienestar del amado es el centro de preocupación del amante. De manera análoga al modo en que Frankfurt excluye el juicio sobre lo valioso del tema de las preocupaciones, el amante ama a su amado no porque vea en él algo especialmente valioso o meritorio, sino porque los intereses o el bienestar de este último son de la mayor importancia para el amante mismo. Y lo son tanto, que traicionar la “autoridad” de las exigencias que nacen de ese amor, exigencias que no tienen, de nuevo, ninguna relación con lo moral, equivaldría para el amante a una verdadera traición a *sí mismo*, una traición en contra de la propia esencia de lo que éste es como persona.¹⁵⁶

¹⁵⁶ Es importante señalar que cuando Frankfurt habla del preocuparse, se refiere a aquello que un agente considera importante para *sí mismo*. Pienso que esto guarda estrecha relación con el hecho de que, igualmente, afirme que traicionar al objeto amado termina por ser una traición *a sí mismo*. De allí que algunos críticos de Frankfurt, especialmente Velleman (2002) y Scanlon (2002), crean que estaríamos ante un cierto narcisismo, que sería consistente, de un lado, con el anticognitivismo en virtud del cual nuestro autor se niega a introducir *lo normativo* en el *caring* y, de otra parte, con su idea de una “esencia” o de una “naturaleza esencial” que, según Frankfurt, nos caracteriza y que habría que “defender” como se defienden esos límites a los que antes nos hemos referido como aquellos que implican una “necesidad” cuya contravención sería “impensable” para el agente. Hacer, entonces, aquello a lo que Frankfurt se refiere como “lo impensable”, para él sería algo muy parecido a “estallar”, a desintegrarse como persona. Esta misma “desintegración” del agente se daría en casos de una grave *ambivalencia*, concebida por Frankfurt como lo opuesto al compromiso claro e *incondicional* de una persona con alguna de sus voliciones de orden superior, esto es, lo contrario a aquello que nuestro autor llama “incondicionalidad” (*wholeheartness*). Al respecto, véase especialmente: 1998b y 2007b. La ambivalencia se presenta, entonces, cuando la persona no puede elegir entre dos deseos incompatibles y que atañen a acciones que tendrían efectos definitivos, nada triviales, para *su vida*. El ejemplo más dramático que propone Frankfurt es el de Agamenón cuando se ve ante la disyuntiva de sacrificar a su hija Ifigenia o, según nuestro autor, perjudicar a su ejército. La interpretación que hace Frankfurt de este ejemplo y, en general, su tesis de que las personas tienen una “esencia” que se puede “romper”, así como su visión negativa del hecho de que a

Creo que la introducción de este tema del amor, en vez de solucionar los problemas que acabamos de señalar, por el contrario, truncan la posibilidad de su solución, con lo cual, la expulsión de lo normativo del ámbito del *caring*, termina por ser un lastre para la plausibilidad de buena parte de la propuesta de Frankfurt. Para ilustrar hasta donde llega el peso de este “lastre”, a continuación precisamente me ocuparé de una de las consecuencias a las que pienso que puede llevarnos: la posibilidad de que todo el complejo mecanismo de la evaluación de los deseos y el logro de la “identificación” con algunos de ellos, finalmente, podría ser visto como un mecanismo inexplicable y sobre el que no cabe dar razones.

3. La “jerarquía” de los deseos y el problema de la “identificación”

Como ya se ha mencionado, Frankfurt afirma que las personas se caracterizan por la necesidad de ejercer una agencia “unificada”, necesidad que viene dictada, de una parte, a causa de la diversidad de sus deseos, que pueden no ser coherentes entre sí y que bien podrían jalonar su voluntad en diversas direcciones. De manera que, piensa nuestro autor, nuestra agencia se desintegraría o quedaría desgarrada por las fuerzas encontradas que ejercerían nuestros múltiples deseos. De allí que, recordemos, Frankfurt piensa que se hace necesario que nos “identifiquemos” sólo con algunos de ellos y, por contra, rechacemos a otros, so pena de no poder hacer de ninguno de nuestros deseos aquel que nos mueva finalmente a actuar, esto es, nuestro deseo “efectivo”. Empero, de otra parte, habría otra dificultad adicional que igualmente haría necesaria nuestra identificación con alguno de nuestros deseos, con miras a ejercer una agencia unificada, y una agencia que produzca acciones y no solamente evaluación de deseos.

Dicha dificultad se debe al peligro de una regresión infinita que podría producirse dado que todo deseo es susceptible de ser evaluado, con lo cual se generaría un metadeseo correspondiente al que, a su turno, también se podría someter a

veces a los agentes humanos nos cuesta decidir entre opciones excluyentes, son objeto de duras críticas por parte de Velleman (2002) quien, repito, también asocia la postura de Frankfurt con un cierto narcisismo. Esta postura de Frankfurt en contra de la ambivalencia, a su vez, se conecta, según Velleman, con otra característica que le atribuye a Frankfurt: un cierto horror (casi neurótico) a “lo otro”, bien sea que se trate de lo otro en uno mismo, bien de lo otro que está por fuera del sujeto, esto es, aquello que amenazaría nuestros “límites”.

evaluación, sin que esta cadena de deseos y metadeseos tuviese, en principio, un tope o un punto final en la dirección que va desde los deseos más básicos hasta los más elevados. Este esquema es calificado por el propio Frankfurt como “jerárquico”, si bien nuestro autor en un principio —en los textos en los que propone inicialmente su idea de la evaluación de los deseos— no parece haber considerado que la mencionada regresión infinita corriera el riesgo de ser insalvable (v.g. 1998), hasta que, en textos posteriores se encuentra de nuevo con dicho problema. Y, sobre todo, se enfrenta a objeciones bastante difíciles de sortear, tales como las que le plantea G. Watson en los años 70s¹⁵⁷ y, más recientemente, a partir de los 90s, aquellas que encontramos en las críticas de autores como M. Bratnam (2002), T. M. Scanlon (2002) y el mismo Watson (2002). En lo que sigue, se intentará mostrar la relación que creo que habría entre: 1) Las dificultades que supone el propio planteamiento inicial del esquema jerárquico (de deseos de orden inferior y metadeseos de orden superior); 2) La forma en que Frankfurt concibe la “identificación” del agente con uno de sus deseos para que sea su deseo efectivo, es decir, la naturaleza de dicha identificación; 3) Su idea acerca de la necesidad del ejercicio de una agencia “unificada”; y 4) La naturaleza que nuestro autor le atribuye a las preocupaciones de la persona que, como ya hemos visto, excluye todo elemento normativo.

3.1 ¿Una cadena infinita de deseos y una decisión inexplicable?

Si, como acabamos de ver, el mecanismo de la evaluación de los deseos bien podría acarrear el problema de que el agente no logre, finalmente, decidirse a actuar, al no poder hacer de uno de sus deseos su deseo efectivo, esto supondría la imposibilidad del ejercicio de una agencia “unificada”¹⁵⁸ y, con ello, del goce de la libertad de la voluntad, la cual, como arriba se comentó, amén de que caracteriza a un agente que logra actuar movido únicamente por aquellos deseos con los cuales se identifica, parece incluir en su mero concepto precisamente la acción, por oposición a la parálisis. Pero si existe el peligro de que la mencionada identificación no ocurra, puesto que, repito, todo

¹⁵⁷ Cf. Watson (1975). Le recuerdo al lector que, si bien la compilación en la que aparece “Freedom of the Will . . .” se publica en los 90s, en ella se incluyen los trabajos que había presentado Frankfurt en los años 70s. De allí la fecha del artículo de Watson.

¹⁵⁸ Remito al lector a la nota de pie de página en la que menciono lo que señala Velleman sobre el tema de la ambivalencia y la “unidad” de agencia, tan deseada por Frankfurt.

deseo puede dejar de ser un deseo efectivo si se lo somete a una evaluación, y dicha evaluación no tendría, en principio, por qué detenerse, entonces Frankfurt cae en cuenta de que esto significaría la amenaza de una quiebra en nuestra capacidad de agencia, o de una parálisis de nuestra voluntad. Con lo cual estaríamos ante la posibilidad misma de dejar de ser agentes o, como el mismo Frankfurt dice, “personas”.

There is no a theoretical limit to the length of the series of higher and higher orders; nothing except common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order. The tendency to generate such a series of acts of forming desires, which would be a case of humanization run wild, also leads toward the destruction of a person (1998, p. 21).

¿Cómo evitar que se cumpla este temor de que la dinámica de la evaluación de los deseos, que en principio debería ser la expresión de esa capacidad reflexiva y autoevaluativa que nos distingue de otras especies, y que nos hace verdaderamente dueños de nuestros actos, termine, paradójicamente, por ser un proceso absurdo y paralizante? ¿Cómo evitar caer en una aporía ante la cual no quedaría otro camino que dar al traste con el mecanismo de la evaluación de los deseos, que parece estar en el corazón mismo de la propuesta de Frankfurt? En “Identification and the wholeheartedness” (pp. 165 y ss.), nuestro autor intenta precisamente salvar dicho mecanismo de las críticas de G. Watson (1975), quien señala que la cadena de deseos y metadeseos es en sí misma inexplicable y arbitraria, del mismo modo en que lo sería la decisión que tendría que tomar el agente que quiera “detener” la cadena, esto es, “parar” de evaluar sus deseos y “elegir” un deseo para hacer de éste el motivo que le impulse a la acción.

Esta decisión del agente requeriría por parte de éste un “acto de voluntad” mediante el cual la persona decida hacer un compromiso definitivo con un deseo que, a su vez, decide no evaluar más. ¿Pero cómo justificar dicha decisión o dicho acto de voluntad? ¿Por qué el deseo con el que el agente, finalmente, se identifica y se compromete no es también sometido a evaluación? ¿Qué es lo que autoriza a “detener” la cadena que, como se ha dicho, en principio bien podría ser infinita? ¿Acaso la cadena misma, y el propio mecanismo que la genera, no serían tan inexplicables y arbitrarios como la decisión de ponerle un límite? Frente a Watson, Frankfurt tratará de mostrar

que la decisión del agente de hacer este compromiso y, de esta manera, de disolver el conflicto entre sus voliciones de segundo orden, no sólo no es arbitraria, sino que, además, es el medio por el cual el sujeto se puede constituir a sí mismo en un “yo” (*self*) unificado, autoconciente y reflexivo (*selfaware, selfreflective*) que asume la responsabilidad sobre esos mismos deseos con los cuales se compromete, y a los que integra dentro su historia como “yo”. Para defender su posición, Frankfurt acude al siguiente ejemplo. Si una persona está llevando a cabo una serie de cálculos aritméticos y en determinado momento del proceso decide detenerse y, por alguna razón, adoptar un resultado, es porque la persona confía en que hay buenas razones para hacerlo. Es decir, confía en que, dado el proceso que ha adelantado, el resultado elegido es el menos incierto y, por tanto, ya no necesita seguir haciendo más cálculos. Es posible que, aunque el sujeto tema que dicho resultado no sea “el” correcto, sin embargo se decida por él porque se percata de los costes que conllevaría el seguir calculando, en lo que se refiere al tiempo, los esfuerzos, el dinero y demás recursos que habría que invertir, los sacrificios que habría que hacer en términos de otros bienes, así como la pérdida de otras oportunidades que se desaprovecharían si se insiste en continuar haciendo cálculos.

Estos costes, pues, pueden ser percibidos por el agente como demasiado altos si se los compara con las eventuales “ganancias” que se obtendrían si el sujeto se empeña en minimizar sus posibilidades de errar. De esta manera, la solución más saludable para la persona será hacer un “compromiso decisivo” con un determinado resultado sobre el que, si bien aún puede abrigar dudas, no obstante representa la salida más razonable frente a la alternativa de seguir calculando, alternativa que, repito, podría ser más costosa que el contentarse con un resultado no del todo seguro. Pues este último puede ser considerado como el menos incierto dentro de lo que ha podido calcularse, amén de que le evitaría al agente incurrir en costes desproporcionados. No obstante, la falta de una absoluta certeza y el haber asumido el resultado elegido como la salida más sensata no significa que el proceso esté cerrado inexorablemente, ya que también cabe la posibilidad de que un agente que se haya visto impelido a elegir un determinado resultado, posteriormente comience a sentirse intranquilo con respecto a éste y, por consiguiente, decida volver a revisarlo si esto mismo no le representa una opción demasiado costosa comparada con la alternativa de dejar las cosas como las había

dejado antes. En este último caso, nos dice Frankfurt, el sujeto podría, y de un modo al que cabe considerar “sensato” o “razonable”, tomar la decisión de volver a revisar los cálculos hechos anteriormente y, por esta misma vía, también revisar el compromiso que antes había adquirido con el resultado sobre el que ahora está comenzando a dudar. El proceso, por tanto, no tiene que ser visto bajo la óptica de la fatalidad como algo necesariamente irreversible, ni destinado al fracaso, entre otras razones, porque tal vez el “éxito” al que se aspire no tenga más remedio que ser relativo o parcial.¹⁵⁹

Esto que ocurre en el caso de quien está llevando a cabo cálculos aritméticos podría servir como modelo, según Frankfurt, de lo que ocurriría tratándose de un sujeto que está tratando de resolver un conflicto entre sus *deseos*. El agente que en determinado momento llega a dudar de un deseo sobre el que antes parecía estar seguro, tiene ante sí la posibilidad de volver a someterlo a evaluación, análogamente a como se podrían revisar unos cálculos que en su momento parecían ser los menos inciertos. En el caso de los deseos, entonces, también se pueden volver a revisar los resultados que anteriormente se habían asumido como los menos inseguros. Ahora bien, igualmente es posible que, estando en ese proceso de reevaluación, la persona nuevamente tome la decisión de “parar”: de no continuar examinando el compromiso adquirido anteriormente. Por ende, puede resolverse a no proseguir con dicho examen y asumir un compromiso decisivo con un determinado deseo, puesto que se ha percatado de que el anterior conflicto que se presentó a nivel de sus deseos, conflicto que le llevó en su momento a dudar y reevaluar, puede que ahora sea percibido bajo otra luz. Es decir, el agente puede que ahora lo vea como un conflicto mucho menos severo de lo que anteriormente creyó, y de allí que la persona se sienta ahora más segura al asumir un compromiso con un determinado deseo.

También existe la posibilidad de que, de modo análogo a lo que ocurría en el anterior ejemplo de los cálculos, al agente le suceda que en un determinado momento caiga en la cuenta de que los costes de seguir deliberando y reevaluando sus deseos se vuelvan demasiado onerosos para su capacidad de agencia y sus posibilidades de acción. Si esto fuera así, la persona podría arribar a la conclusión de que lo más sensato es que tome una decisión y “pare” de evaluar, para así asumir un compromiso con un

¹⁵⁹ En este sentido, cabría recordar la noción de “racionalidad imperfecta” de J. Elster (1995, pp. 66-188), así como la crítica de N. Rescher (1993, pp. 64-89) a cierta forma de escepticismo que se apoyaría, según él, en un señalamiento de los límites de la razón.

deseo sobre el cual, si bien aún abrigaría dudas, no obstante dichas dudas serían mucho más llevaderas que los costes de un largo proceso de evaluación, proceso que acaso no parezca prometer mayores recompensas. Esta decisión del agente no tiene, por tanto, que ser calificada de “arbitraria” o “absurda”, pues, como se ha visto, habría buenas razones para revisar los compromisos sobre los que abrigamos dudas preocupantes y que significan un conflicto grave para nuestra identidad o para nuestra unidad. Pero también habría buenas razones en el caso de una persona que asuma que ya es hora de hacer un compromiso y detener su proceso de evaluación, puesto que el sujeto percibe que el conflicto que anteriormente le llevó a dudar y evaluar, ahora ya no significa más un conflicto que amenace con desgarrar su propia identidad. Igualmente, habría buenas razones que asistirían al agente que se dé cuenta de que su compromiso con un determinado deseo, por más que éste no represente la opción más segura es, sin embargo, la salida menos costosa para su identidad, e incluso para el ejercicio de su capacidad de agencia; mientras que, por el contrario, dicha capacidad se vería seriamente amenazada si la persona opta por continuar con el costoso proceso de seguir evaluando. Lo cual no sería lo más razonable si el agente advierte que el alargamiento del proceso de evaluación no le promete más beneficios que aquellos que, por lo pronto, en cambio sí le promete su compromiso con un deseo que ponga término al proceso, y que permita que se produzcan, por fin, la decisión y la acción.

Pienso que el anterior argumento de Frankfurt resulta bastante ingenioso y tiene la ventaja adicional de que nos llama a la sensatez. Empero, también creo que la polémica con Watson nos hace caer en la cuenta de que aún no tenemos del todo clara la justificación ni la naturaleza de la evaluación de los deseos. ¿Por qué razón el agente decide evaluar sus deseos? ¿Por qué a algunos de esos deseos llega a considerarlos como “deseables” y a otros no? ¿Por qué se compromete con un deseo a partir de ese juicio sobre lo deseable/indeseable? ¿Qué justifica, qué hace “razonable”, dicho juicio, así como el propio mecanismo de la evaluación de los deseos? ¿Cómo estar seguros de que esta jerarquía que se plantea entre deseos y metadeseos no es, en último término, arbitraria? ¿Cómo no considerar inexplicable y arbitrario el acto de voluntad por el cual el agente decide dejar de evaluar sus deseos y, entonces, opta por uno de ellos? En mi opinión, para responder a estas cuestiones Frankfurt bien podría señalar en dirección a aquello que parecer ser postulado como el núcleo de la propia identidad del agente, eso

que le permite a la persona “filtrar” y “jerarquizar” sus deseos: sus preocupaciones fundamentales. Creo que si nuestro autor permitiera que se “abriese” el contenido de estas últimas, si se examinara más de cerca en qué consisten y a qué obedecen, tal vez se podría mostrar mejor qué sentido tiene el proceso de evaluación de los deseos y por qué éste no tiene necesariamente que ser pensado como un regreso infinito que, en último término, resultaría, como acierta en señalar Watson, inexplicable o absurdo.

No obstante, también me parece que la articulación de una respuesta convincente a críticas como las de Watson podría encontrarse con el obstáculo de que, como arriba se ha señalado, en Frankfurt no puede apreciarse una exposición clara sobre la naturaleza de las preocupaciones y, lo que resulta más difícil de aceptar: excluye de éstas todo elemento normativo. Es aquí donde creo que reside el problema de que no logre explicarse el mecanismo mismo de filtrar los deseos, ni pueda mostrarse por qué dicho mecanismo implica necesariamente una jerarquía. Tampoco se explicaría cómo puede darse cuenta, en último término, de los metadeseos más “elevados” de dicha jerarquía —habría que suponer que tendrían que ser las preocupaciones “más” fundamentales—, ni del compromiso que con tales metadeseos tenga el agente. De este modo, permanece sin aclarar cómo pueden ser pensadas las relaciones que pueda haber entre las preocupaciones fundamentales, o entre ellas y las otras posibles fuentes de criterios que tenga el agente para filtrar sus deseos. A lo cual habría que agregar el hecho de que en sus textos más recientes nuestro autor termina por explicar las preocupaciones en unos términos que, como bien señalan no sólo el propio Watson, sino unos cuantos autores más, impedirían verlas como una fuente de criterios justificables para las decisiones de las personas. Así, este modo de mostrar la naturaleza de las preocupaciones (por parte de nuestro autor) no permitiría que éstas contaran como criterios discutibles más allá de la vivencia inmediata y la fuerza inarticulada con que pueda experimentarlas el propio sujeto quien, por tanto, tampoco podría aclarárselas a sí mismo. Concretamente, me refiero a una tesis de Frankfurt que arriba ya se comentó: la tesis de que aquello que estaría en el centro mismo de tales preocupaciones es el “amor” (2007a), entendiéndose por este “amor” algo ajeno a la posibilidad del dar razones. Con lo cual, estaríamos aún más lejos de lo que tan prometedoramente nuestro autor parecía anunciar con su idea de la evaluación de los deseos y de la capacidad reflexiva que ésta revela.

Por lo anterior, pienso que críticas parecidas a las de Watson posiblemente seguirían incomodando a quien quiera evitar que el relato de Frankfurt conduzca a una aporía y tengamos que resignarnos a pensar que la complejidad volitiva de las personas, y la dinámica de la evaluación de los deseos parecerían descansar en meros actos de voluntad sobre los cuales no cabe dar razones. En otros términos, como lo demuestran críticas del tenor de las de Watson, también podría hacerse una lectura del relato de Frankfurt según la cual los deseos, tal y como él los muestra, en último término seguirían sin poder explicarse, permanecerían como hechos brutos que el agente “decide” aceptar o no aceptar. Lo cual, además, se agrava si se supone que dicha “aceptación” se funda sobre la base de unas preocupaciones y de un amor que no se muestran de una forma comparativamente más compleja que aquella bajo la cual se presentan esos *data* inexplicables que, como vimos, parecían incomodar a Frankfurt en su crítica a la moderna teoría de la acción.

Creo que lo anterior nos hace ver la importancia de que se incluya lo normativo en el cuadro que Frankfurt presenta de la formación de las voliciones de segundo orden, la cual sería lo que distinguiría a una persona de un *wanton*. Sólo si lo normativo juega algún papel en dicha formación, entonces podría entenderse que un agente se comprometa con una de sus voliciones de segundo orden y haga de ella su deseo efectivo, poniendo fin a esa cadena, en principio infinita, de evaluación de sus deseos. Así, el compromiso definitivo del agente con un determinado deseo, esto es, su “identificación” con dicho deseo para que éste sea su deseo “efectivo” y le mueva a actuar, podría salvarse de ser visto como un acto de voluntad completamente inexplicable, pero sólo a condición de que el sujeto pueda explicarnos, apelando a razones, tal acto de decisión. Si lo normativo y las razones no cumplen ningún papel en las preocupaciones de las personas, entonces, casos como el del adicto preocupado que lucha contra su adicción no se podrían explicar ni siquiera en los propios términos de Frankfurt, ya que en dichos casos no se verían ejemplos de una “evaluación” de los deseos ni de una preocupación por una identidad ni, en general, de aquello que permita hablar de una diferencia entre personas reflexivas y autocríticas, frente a *wantons* no concernidos por la “deseabilidad” de sus deseos.

Si la actitud del adicto contrariado se nos presenta como radicalmente distinta a la del adicto despreocupado es porque, precisamente, el primero tendría algún tipo de

razón para querer cambiar sus deseos y para ver a algunos de ellos como “indeseables”. Si no fuera así, ¿qué podría, entonces, hacerlos aparecer ante sus ojos como indeseables? Independientemente de si el adicto preocupado apela o no a razones estrictamente morales, o de otro tipo, v.g. religiosas, o del tenor que sean, en todo caso no creo que su actitud se logre explicar sin suponer que haya tales *razones* de por medio –repito, con independencia de la naturaleza específica de esas razones–. Por esta misma vía, tampoco podríamos entender dicha actitud si no interviene alguna forma de explicación –así sea sólo *in foro interno*– por parte del agente que la manifiesta, explicación en la que debe apelarse a algún tipo de referente normativo.¹⁶⁰ Sin este último, entonces, pienso que nada evitaría que la jerarquía de deseos se torne en una regresión infinita, ya que en ella no podrían jugar su importante papel las preocupaciones de las personas, puesto que éstas no proporcionarían el contexto articulable y justificador de las voliciones y acciones humanas.

Así, la jerarquía de deseos, y el acto de decisión que evite que ésta se torne en una cadena infinita, se salvarían de la irracionalidad y del capricho únicamente si cabe la posibilidad de que dicho acto se muestre como haciendo parte de un contexto en el que adquiere sentido; contexto que, vistos los aportes de Frankfurt, creo que sólo podría ser proporcionado por aquello que él llama “preocupaciones”. Pero, para que este contexto deje de ser un misterio, sería necesario que las preocupaciones se “abran”, dejen de ser opacas e inanalizables, podamos ver su contenido, podamos hablar de él. Sólo así el acto de identificación con un deseo dejaría de ser un salto en el vacío o una suerte de acto de fe. En la siguiente sección, intentaré mostrar una posible manera de lograr que, a pesar de los fallos señalados hasta ahora en el esquema de Frankfurt, se logren salvar aquellos aportes suyos que creo que nos pueden llevar más allá de Gauthier, pudiendo, por tanto, dichos aportes significar una contribución al objetivo de esbozar un modo de agencia humana que resulte más complejo y plausible que aquel que pudimos obtener de *La moral por acuerdo*.

¹⁶⁰ Sin embargo, nuestro autor, en sus últimos textos, sigue siendo enfático en negar que en el acto de identificación con un deseo esté involucrada una deliberación en la que intervengan “razones”. Lo cual se ve claramente en sus respuestas (2002, 2002c y 2002f) a las críticas que ha recibido por parte del propio Watson (2002) y de autores como Bratnam (2002) y Scanlon (2002).

3.2 Una propuesta para fortalecer el agente “integrado” de Frankfurt, en contraste con el maximizador de Gauthier

Como de algún modo ya se ha mencionado, nuestro autor expresamente atribuye una concepción “fragmentada” del agente humano a cierto modelo de explicación de las acciones a partir de unos deseos y creencias a los que, según Frankfurt, se los ha asumido como eventos que meramente “les suceden” a las personas. Como arriba se vio, según este modelo del que quiere desmarcarse el autor, a los agentes humanos se los pensaría básicamente como seres a los que “les pasa” que van teniendo determinados deseos y creencias a lo largo del tiempo, de manera análoga a como “les pasa”, por ejemplo, que van sufriendo accesos de tos o ataques de sueño, se les van ocurriendo ideas o son asaltados por sus recuerdos. El resultado de ver de este modo “fragmentado” a las acciones, creencias, deseos y demás eventos mentales es que, según Frankfurt, entre ellos no se aprecie una conexión, por más que ocurran de manera sucesiva; no aparecen dentro de lo que nuestro autor llama una “historia integrada” (1998d, pp. 80-94).¹⁶¹ A su vez, el agente mismo aparece como desintegrado en la multiplicidad de los deseos y creencias que le van sobreviniendo a lo largo del tiempo, y de las acciones que va emprendiendo a causa de esos deseos y creencias, según la

¹⁶¹ Sobre esta noción de “historia integrada”, creo que en nuestro autor se puede advertir una suerte de cambio entre sus textos de los años 70s-80s, y aquellos otros que publicó a partir de los 90s. En los primeros le concede un significado especial a esta idea de que los seres humanos podemos pensarnos a nosotros mismos como teniendo una “vida” y una “historia”, mientras que, en mi opinión, en los segundos pareciera que Frankfurt le resta importancia al hecho de que podamos señalar esa historia como parte de las *razones* para nuestro *caring*. Ello tal vez se relacione con el hecho, ya señalado arriba, de que también en esta “segunda época” el autor es más explícito en negar el peso que tendría el apelar a “razones” cuando se trata de dar cuenta de nuestro *caring* y del compromiso que asumimos con algunos de nuestros deseos. Esto último, a su vez, creo que podría explicar que, a pesar de la preocupación que Frankfurt manifiesta sobre la necesidad de una “unificación” de la agencia humana, por la vía de evitar experiencias tales como la ambivalencia (opuesta a la “incondicionalidad” o *wholeheartness*), y exagerando, tal vez, según críticos como Velleman (2002), el peligro de la ambivalencia, y postulando como modélico un yo absolutamente unificado al que le corresponde una “esencia”, sin embargo, nuestro autor en algunos apartes parece que le resta importancia a la apelación que pueda hacer un agente a su “historia” al momento de asumir un compromiso definitivo con una de sus voliciones de orden superior; es decir, cuando se produce la “identificación” que evita la ambivalencia. Este posible cambio en cuanto a la relevancia atribuible a la historia del agente, incluso lleva a Frankfurt, en sus últimos artículos (especialmente 2007a), a decir que algunos de nuestros deseos son el producto de nuestra *mera* voluntad, sin que en ella quepa algún tipo de proceso deliberativo, y sin que tengamos que involucrar en este asunto una apelación a nuestra “historia”.

formulación que, como se comentó, Frankfurt le atribuye a la moderna teoría de la acción. Desde el modelo que nuestro autor asocia con dicha teoría, el agente y su vida surgen, entonces, como colecciones de fragmentos, no como unidades integradas y, en contra de esta concepción, Frankfurt intenta hacer plausible su idea de que las personas y sus acciones podrían ser asumidas como “integradas” en torno, precisamente, de aquello que nuestro autor llama “preocupaciones”.

Desires and beliefs can occur in a life that consists merely of a succession of separate moments, none of which the subject recognizes –either when it occurs or in anticipation or in memory– as an element integrated with others in his own continuing history. When this recognition is entirely absent, there is no continuing subject. The lives of some animals are presumably like that. The moments in a life of a person who cares about something, however are not merely inherently by formal relations of sequentiality (1998d, p. 83).

Llegados a este punto, creo que podríamos ensayar la hipótesis de no dar al traste con los aportes de Frankfurt e intentar ver su relevancia para una noción menos restringida de la agencia humana, para lo cual puede resultar útil esbozar una suerte de paralelo con el relato que obtuvimos de Gauthier. Pienso que con lo que llevamos hasta aquí, ya contamos con algunos elementos que nos permitirían afirmar que, si asumimos la perspectiva de Frankfurt, tal vez estaríamos más cerca de lanzar la mirada que tendría el propio agente sobre sus acciones y aquello que le motiva a realizarlas, mientras que, si nos situamos en el punto de vista del autor canadiense, acaso tendríamos que contentarnos con lanzar una mirada más propia de un *observador* que *no participa* en tales acciones,¹⁶² sino que meramente las describe e intenta explicarlas causalmente, señalando como base de dicha explicación unos deseos y unas creencias atribuidas al agente que es objeto de observación. Lo cual encaja con los supuestos conductistas que, según se indicó, parecen estar presentes en Gauthier. A su vez, estos deseos y creencias a los que se asume desde un modelo como el que suscribiría el autor canadiense, igualmente serían vistos desde “fuera” y por tanto, serían considerados como *data*, lo cual contrasta con unos deseos que pueden ser objeto de vivencia, exploración y, sobre todo, de *evaluación* y crítica por parte de un agente como el que aparece en el relato de

¹⁶² En este punto, le recuerdo al lector la referencia que, al final del tercer capítulo, se hizo al llamado de atención que hacen autores como Strawson (1995) y Habermas (1989 y 1994a), para que el filósofo moral asuma la perspectiva del agente que *participa* en la praxis de la que intenta dar cuenta, en contraste con la actitud del *observador* que “objetiva” dicha praxis.

Frankfurt. Este último tipo de agente, por tanto, no asumiría sus deseos como meros *hechos* que no tiene más remedio que aceptar, sufrir, reprimir ni, mucho menos, limitarse a describir al modo en que se describen los hechos externos a uno mismo.

Obviamente, quien se limita a señalar tales *data* lograría, repito, ofrecer explicaciones de tipo causal, las cuales sin duda pueden tener un lugar legítimo e incluso necesario. Pero el caso es que, si el análisis de los deseos y las decisiones humanas se limitara a lo que pueda decirnos este tipo de observador, y a esa mirada suya que, parafraseando la expresión de Habermas (1989 y 1994a), “congela” o convierte al otro, a uno mismo y al mundo en meros “objetos”, entonces quedaría oculto lo que, en cambio, sí vivencia quien *participa* en la acción y trata de dar cuenta de ella desentrañado, ya no sólo sus causas, sino también su *sentido*. Esto es, el agente que intenta reconstruir dicho sentido y debe habérselas *–in foro interno–* con la complejidad de sus deseos. Creo que, si se la compara con lo que vimos en Gauthier, lo que resulta interesante de la propuesta de Frankfurt es que un agente que intente dar cuenta de sus acciones y deseos contaría para ello con más posibilidades que simplemente asumirlos como meros eventos que “le suceden” y que, como tales, indudablemente tendrían causas y producirían efectos eventualmente visibles para otros agentes. El agente dibujado por Frankfurt también podría, y tal vez con un sentido de mayor prioridad o de urgencia para ella/él, experimentar sus deseos como un entramado cuyos significados, relaciones de prioridad, conexiones conceptuales y valores ella/él necesita *–al menos de tanto en tanto–* aclararse para sí misma(o).¹⁶³ Lo cual, a su vez, tendría relación con otro importante punto de contraste entre Gauthier y Frankfurt: la mención que este último hace de una “historia” y de una “vida”, las cuales creo que designarían unidades de sentido mucho más complejas, y a partir de las que se podría obtener una mayor riqueza de recursos para el análisis de las acciones humanas.

Tales recursos me temo que difícilmente podrían ser hallados en la propuesta de Gauthier, sobre todo si tenemos en cuenta la centralidad que en ella tienen su figura de un maximizador autointeresado y su noción de “preferencia”. Creo que, en una clara diferencia con Frankfurt, a partir de dicho esquema se contaría con menos elementos

¹⁶³ Hago uso de este vocabulario con el que posiblemente Frankfurt no se encontraría del todo a gusto, pues creo que le permite al lector ver con más claridad la idea que intento resaltar: la mayor complejidad que creo que caracteriza, tanto al agente pensado por Frankfurt, como a los deseos de dicho agente, si se compara al primero con el tipo de agente modélico con el que se compromete el discurso de Gauthier.

para que el sujeto pudiera abordarse a sí mismo y, al mismo tiempo, para que otros agentes, e incluso quienes estemos interesados en una aproximación meramente conceptual o filosófica, podamos acercarnos un poco más al *contexto* y al trasfondo sobre el que cobrarían *sentido* las preocupaciones, deseos y acciones del agente. Expresiones como “historia integrada” o “vida”, si bien no dejan de ser vulnerables a la crítica y a la sospecha, creo que tienen la ventaja de ser más útiles para quien se empeñe en dar cuenta de tales preocupaciones y acciones; sobre todo para quien esté interesado en el significado que los agentes, asumiendo más el papel de *participantes* que de *observadores*, quisieran desentrañar de sus propias acciones y de la “vida” (y de la “historia”) en la que dichas acciones cobrarían un sentido.

Buena parte de esto último tal vez se le escape a quien se contente simplemente con “describir” los hechos, incluso tratándose de hechos supuestamente “internos”, tales como las creencias y los deseos cuando, repito, se los asume únicamente como *data* inanalizables. Y ello sin mencionar el peligro que podría presentarse de que la “descripción” que podría ofrecer un observador como el que tendría en mente Gauthier, pudiese, a su vez, limitarse a cumplir con los dictámenes de la asepsia valorativa que exige el uso de un lenguaje puramente observacional, lenguaje que igualmente estaría más cerca de obedecer al ideal reduccionista con el que, según vimos, podríamos emparentar al filósofo canadiense. Dicho lenguaje, de ser utilizado en una descripción causal influiría en que ésta no fuese lo suficientemente profunda, nos dijera muy poco, no alcanzara a ser una propiamente una “explicación”, al no permitir que se diera cuenta de algo más que la mera conducta observable de elección de unos agentes que son vistos desde fuera, y en cuyas profundidades no se entraría, si se aborda dicha conducta únicamente echando mano de categorías como las que asume Gauthier, sobre todo, su noción de “preferencia”.

Por lo anterior, pienso que, comparado con el análisis hecho por este último, el análisis que hace Frankfurt resulta bastante revelador, puesto que desde él puede mostrarse, por lo menos en mayor medida, eso que queda oculto si se asume un esquema como el que se aprecia en el discurso del filósofo canadiense. Incluso, creo que puede utilizarse dicho análisis dentro de una estrategia más amplia, orientada a mostrar qué es aquello que tal vez no se nos pueda decir acerca de la agencia y de las motivaciones humanas si éstas se ven sólo “desde fuera” y como meros *data*

incuestionables u opacos, reducidos a la base motivacional única, y también opaca –y también “reducida”– del *autointerés* que caracteriza al *agente maximizador* de Gauthier; un autointerés que, por tanto, tampoco sería susceptible de mayores análisis y al que se supone como reflejado o decantado en las *preferencias* atribuibles a dicho agente. Igualmente, la visión que tenga el observador de tales preferencias alcanzaría para hacer una valoración que llegaría a límites como el del juicio sobre la “coherencia” de la *conducta* de elección del sujeto que es observado. En mi opinión, si se asume el punto de vista de un observador provisto de las herramientas sugeridas por Gauthier, se haría un tanto difícil superar dicho límite, no se lograría dar cuenta de buena parte de los detalles que nos acercarían a un relato un poco más completo sobre el sentido de una acción, más exactamente, sobre lo que significa para el agente que participa en ella y que tal vez no pueda menos que verla, o intentar verla, en conexión con su “vida” o en el contexto de su “historia”. De allí que crea que, si contrastamos lo que llevamos visto de la propuesta de Frankfurt, con lo que se señaló en la de Gauthier, acercarnos a la primera y alejarnos de la segunda posiblemente nos haga también más fácil acercarnos a lo que nos respondería un agente a quien le preguntáramos sobre el sentido de su acción; incluso aceptando, por supuesto, la ineludible falibilidad del propio agente, que bien podría cometer errores o autoengañarse al ofrecer su relato acerca de sus acciones y motivaciones.

Dejando por lo pronto de lado el peligro del autoengaño, sobre el que se volverá en el siguiente capítulo y, por supuesto, reconociendo el límite que se alcanzó al haber Frankfurt excluido cualquier elemento normativo de las preocupaciones de los agentes, sin embargo, creo que vale la pena intentar quedarnos con la pregunta sobre qué comportarían dichas preocupaciones, y qué veríamos en ellas si las liberamos de la opacidad causada por las reticencias de Frankfurt en contra de temas como la moral, la ética, los valores, la verdad en un sentido epistémico y, en general, lo normativo. Si insistimos en “abrir” este cerrojo en el que persiste Frankfurt, podríamos entonces atrevernos a mirar qué es lo que comportan las preocupaciones de las personas, y con ello, sacar mejor partido los aportes de este autor, en especial de su idea de que los agentes humanos también pueden evaluar sus deseos y tomar una distancia crítica frente a éstos. Podríamos, pues, apartándonos de una postura que nos lleve al punto muerto de afirmar que, si deseamos algo, es únicamente “porque así nos apetece y punto”. Lo

interesante de Frankfurt, y por contra de una postura de este tipo, es que permite descubrir que por lo menos algunos de nuestros deseos se deben a que estamos radicalmente *preocupados* por ciertos asuntos a los que consideramos de la mayor importancia. Si nos tomamos en serio esta sugerencia de Frankfurt, pienso que su idea del *caring* nos permitiría abrir la coraza que aparentemente cubre las preferencias de los agentes humanos cuando a éstas se las piensa al modo en que aparecen en un modelo como el de Gauthier. Si dicha coraza es abierta, entonces podría explorarse esa base motivacional que es asumida como incuestionable por parte de las tradiciones con las que se identifica el autor canadiense, lo cual conllevaría un enriquecimiento de nuestra idea del agente humano y de sus motivaciones.

Por eso pienso que no vale la pena insistir en una lectura de Frankfurt en la que nos veamos abocados a concluir que el mecanismo de evaluación de los deseos es inexplicable o es, en último término, irracional. Creo que antes debe más bien ensayarse la hipótesis de que, si se presenta dicho mecanismo, es porque los deseos humanos no son meros hechos brutos, sino que, por más que Frankfurt se empeñe en negarlo, dichos deseos se entrelazan con elementos *normativos* que deben poder ser objeto de articulación, aclaración y discusión. Pero esto último significa que habría que romper con una idea que pienso que ata a Frankfurt a un modelo como el de Gauthier, y que es la causante de las aporías e inconsistencias arriba señaladas: su compromiso con una noción de sujeto solitario, sin vínculos con un contexto, sin nexos con otros agentes, en síntesis, un ser bastante parecido al Robinson que Gauthier postula como su agente ideal. Sólo si se rompe con este compromiso, creo que puede verse por qué tendría sentido evaluar los deseos y por qué dicha evaluación no tiene que conducirnos al vértigo de lo absurdo. Pues en ella el agente apela a referentes normativos que el propio sujeto no se inventa de la nada, sino que comparte o discute con *otros* agentes. Por esto no creo que estemos inexorablemente empujados a sacar la conclusión de que los metadeseos que parecen estar “más arriba” en la jerarquía propuesta por Frankfurt, sean “inexplicables”; ya que estos metadeseos también pueden ser objeto de “lectura”, de análisis por parte de los agentes que se sientan comprometidos con ellos por motivos que, por lo menos en principio, también serían compartibles y discutibles con otros agentes.

Como puede verse, la figura del agente solitario, común a Frankfurt y a Gauthier, no sólo se hace demasiado débil, sino que puede constituirse en un obstáculo que impediría dar cuenta de la naturaleza de nuestras motivaciones y de la manera como leemos y evaluamos tales motivaciones. Por esta razón, pienso que dicha “lectura” o “evaluación” no tiene que ser asociada a un esquema “vertical” de “jerarquía”. Más bien, creo que el compromiso que asumimos con ciertas motivaciones y con ciertos referentes normativos, es algo que cobra sentido, no en una cadena jerárquica que nos puede llevar al absurdo, sino en el contexto de un entramado de significados, de una suerte de *red*. Entender el porqué de un determinado compromiso, requeriría no apelar a un metadeseo que, a su vez, termine por ser inexplicable, sino señalar las *conexiones* de ese compromiso con otros compromisos que hagan parte del mismo contexto dador de sentido, de la misma red, de modo análogo a como, según Wittgenstein, entender una palabra y saber utilizarla requiere que la relacionemos con otras palabras en el contexto de un juego de lenguaje.¹⁶⁴ De allí que piense que, si se insiste en la idea de una jerarquía de deseos puede hacerse más difícil dar cuenta de éstos, hasta el punto en que tengamos que darle la razón a Watson: los metadeseos más “superiores” no se entienden, son objeto de elección arbitraria. Creo que para dar cuenta del sentido de un metadeseo a partir de la cual se filtren otros deseos, no se lo debe insertar en una escala jerárquica –vertical–, sino ponerlo en una relación –horizontal– con otros metadeseos con los cuales conforma una red, un entramado complejo y holista que, a diferencia de una cadena jerárquica, no requiere de un “último” anclaje o fundamento, el cual, a su vez, al no admitir fundamentación alguna, nos dejaría, como bien señala Watson, ante las puertas de lo inexplicable. Pero no es esto lo que resulta de intentar dar cuenta del sentido de un deseo si le vemos como conectado con otros en una red; en este caso, buscar el sentido sería algo parecido a hallar conexiones, a deshebrar los hilos de la red, no a buscar fundamentos infundamentales. De allí que arriba señale una de las ideas que el propio Frankfurt propone y que acaso evitaría que su discurso acabe siendo vulnerable a críticas como la de Watson: tal vez se entienda mejor a las preocupaciones y a los deseos humanos si se los piensa como haciendo parte de una “vida integrada”.

Por lo anterior, en el capítulo siguiente intentaré dar respuesta a los interrogantes ante los cuales parece dejarnos Frankfurt. Buscaré aprovechar los aportes que pueden

¹⁶⁴ *Investigaciones Filosóficas*, Parte 1, §§ 23-50.

hallarse en Charles Taylor al objeto de mostrar cómo, si los agentes humanos están interesados en lo que sus acciones y motivaciones significan para ellos y para esas personas con las cuales dialogan sobre el sentido de tales acciones y motivaciones, entonces se puede poner bajo una luz más favorable las ideas de Frankfurt acerca de nuestra capacidad para evaluar nuestros deseos. Pues estos últimos pueden ser “entendidos”, pueden dejar de ser vistos como *data inexplicables* si, como muestra Taylor, la evaluación que hacen las personas de sus deseos y motivaciones supone que los sujetos pueden aclarar y revisar aquello que Taylor, utilizando un lenguaje prohibido por Frankfurt, define como “compromisos valorativos”. Pero esta salida significa que para contestar a las airadas preguntas de Watson, el *caring* del que habla Frankfurt tendría que hacer parte de un horizonte valorativo que los agentes humanos suponen en sus juicios y preferencias.

Por una parte, la naturaleza compleja y holista de este horizonte, así como su papel de trasfondo, implican que los sujetos no tienen necesariamente claros sus compromisos valorativos fundamentales, esos que están a la base de tales juicios y preferencias y, por otra parte, la necesidad que en ocasiones tienen los agentes de llevar a cabo una labor de aclaración y revisión de esos compromisos valorativos, muestra que dicha labor puede ser pensada como un diálogo con otros. Esto es, con aquellos con los cuales se comparten y discuten lo que las cosas significan para el agente. De tal manera que, desde Taylor, se muestra que no tendría sentido pensar este proceso de aclaración, que es también autoaclaratorio, como la búsqueda del fundamento-del-fundamento, ni como ese ascenso *ad infinitum* al que parece temerle Frankfurt cuando advierte las perplejidades que suscita esa supuesta estructura jerárquica de los deseos sobre deseos. Antes bien, desde Taylor, a los deseos se los puede entender suponiendo el trasfondo de una red de significados, de un horizonte de naturaleza holista contra el cual el agente contrasta sus preferencias y con el cual tiene vínculos, compromisos.

De allí que proponga entender la dinámica de esa estructura volitiva compleja que presenta Frankfurt, como la dinámica que evidencia la manera como involucramos nuestros compromisos valorativos a la hora de determinar qué debemos hacer en un sentido que también (y a despecho de Frankfurt) puede ser moral. Esto es, la preocupación por movernos según aquellas motivaciones que, al decir de Frankfurt, quisiéramos que fueran las que nos movieran a actuar, es una preocupación que

evidencia la forma en que influyen en nosotros esos compromisos valorativos de los que habla Taylor. De hecho, si Frankfurt advierte que una persona es aquel tipo de agente para el cual su voluntad (*will*) es un problema; es alguien que está preocupado por la calidad de las motivaciones que quisieran que lo movieran a actuar (1998, p. 19), podríamos entonces decir, con Taylor: es alguien para quien su identidad es un problema. Pues esta identidad se puede aclarar, “ofuscar” u oscurecer, ganar o perder en luchas más o menos parecidas –aunque, con algo de suerte, no siempre tan angustiosas– a la que tiene el adicto contrariado. Así, la preocupación que distingue a las personas de los *wantons*, creo que, si bien no es de naturaleza exclusivamente moral, por lo menos sí evidencia nuestros vínculos con una identidad y con unos compromisos valorativos que también incluyen lo moral. Por ello, en el siguiente capítulo intentaré mostrar cómo, si se acude al discurso de Charles Taylor, esta pintura de Frankfurt se completa sobre la base de un marco más amplio en el que las cuestiones valorativas, incluyendo las de naturaleza moral, cobran una importancia que no se aprecia tan claramente en la descripción de Frankfurt. Adicionalmente, estas cuestiones valorativas explican buena parte de la dinámica y de la complejidad que Frankfurt le atribuye a nuestra estructura volitiva como personas.

Capítulo 5

El “evaluador fuerte” de Taylor

En este capítulo me propongo utilizar los conceptos de “horizonte valorativo” y de “evaluador fuerte” desarrollados por Charles Taylor, con el objeto de mostrar de qué manera, una vez vista la plausibilidad y la importancia de tales nociones, podría entonces hacerse más clara una idea no restringida de la agencia humana, idea que creo nos permitiría ir mucho más allá del modelo de agente que, a mi entender, surge de una propuesta como la de Gauthier. Dentro de la noción más compleja o no reducida del agente, tal y como intentaré que se aprecie en el discurso de Taylor, igualmente buscaré que se muestre cómo podrían ser incluidos aquellos aportes de Frankfurt que, según lo visto en el capítulo anterior, pienso que sigue valiendo la pena conservar y desarrollar, sin que se tenga que pagar por ello el precio de arribar a los resultados aporéticos a los que parecía llegarse en el transcurso de su exposición. Así mismo, trataré de obtener de Taylor una clave importante para la ruptura con aquello que en Frankfurt se consideró como la causa de los mencionados resultados aporéticos: a saber, los remanentes de un compromiso, aún bastante presente en dicho autor, con una idea del agente humano que a la larga, como intentó mostrarse al final del cuarto capítulo, resultaba bastante similar a la figura del Robinson solitario que Gauthier asume como modelo de agencia humana. En este quinto capítulo me propongo, entonces, mostrar la plausibilidad de una alternativa a dicha figura, a saber, el modelo del “evaluador fuerte” propuesto por Ch. Taylor, como una vía que nos permita escapar de los resultados indeseables a los que nos condujo el hecho de que Frankfurt se negara a incluir referentes *normativos* dentro de su esquema de la evaluación de los deseos y, por ende, dentro de su concepción del *caring* como la clave de lo que él veía como una forma de agencia “integrada”.

Pienso que dicha vía se hace más clara si examinan las implicaciones de una tesis fundamental dentro de la propuesta de Taylor: los agentes humanos deliberamos, hacemos juicios y tomamos decisiones sobre la imprescindible base que nos proporciona el horizonte de sentido, el horizonte referencial que está conformado por nuestros compromisos valorativos más importantes. Dentro de tales referentes de valor y de significatividad, si bien Taylor hace un énfasis especial en el aspecto ético, creo

que tendrían que incluirse, al menos, también aquellos que estarían asociados a lo que un agente considera que se adecua a la verdad y al rigor en un sentido epistémico, a lo justo o correcto en sentido moral, a lo bello en un sentido estético y, por ende, no solamente a lo “bueno” o “valioso” en sentido ético.¹⁶⁵ Lo cual pienso que nos alejaría de las reticencias de Frankfurt frente a lo que este autor vería como ingerencias indebidas, por parte de la epistemología y de la moral, en su concepción del *caring*. Al mismo tiempo, creo que también ofrecería la ventaja comparativa de acercarnos a una idea que Frankfurt comparte pero, en mi opinión, no logra desarrollar del todo a causa de sus prevenciones frente a lo normativo: la noción de una vida humana y de un agente “integrados”. En busca del desarrollo de tal idea, una vez se ponga de manifiesto tanto el papel determinante que juega el horizonte postulado por Taylor, como los innegables nexos que, según él, todo agente humano guarda con dicho horizonte, espero entonces poder mostrar cómo se hace plausible la especial forma de agencia humana que resultaría impensable sin tales nexos, y que Taylor sintetiza en su figura del “evaluador fuerte”. Dicha figura proporciona, a mi entender, un modelo mucho más complejo y convincente que el que podría apreciarse en Frankfurt pero, sobre todo, en Gauthier. De allí que buena parte del presente capítulo se oriente a mostrar las razones que apoyarían esta hipótesis.

Lo anterior no implica que sugiera una ruptura radical entre Taylor y Frankfurt. Más bien intentaré que se revele de qué manera, partiendo de los aportes del primero, podría arrojarse una luz bajo la cual se vieran de manera más favorable aquellos elementos que considero importantes y rescatables de Frankfurt, con la consecuencia de que pueda apreciarse por qué merecería la pena que se siga insistiendo en desarrollar su noción del *caring* y, con ella, su idea de la evaluación que los agentes humanos pueden hacer de sus deseos y preferencias. Por consiguiente buscaré que se vea la importancia que reviste este aporte de Frankfurt, pero de tal forma que, una vez solucionado el escollo que suponía su prohibición de incluir en dicha evaluación elementos normativos, pueda entonces “abrirse” al análisis y a la crítica la base motivacional atribuible a una persona. De este modo, pienso que tal base motivacional perdería la

¹⁶⁵ Recordemos la lección de Habermas (1989 y 1994a) sobre las diversas, aunque conexas, pretensiones de validez que, a partir de los aportes de Austin y Searle, se consideran como supuestas y exigidas tácitamente por parte todo hablante cuando se involucra en un intercambio lingüísticamente mediado.

opacidad que parecía blindarle frente a todo intento de examen, haciéndole inanalizable o indiscutible en el contexto de discursos que, como el de Gauthier, terminan por hacer inviable tal examen o por trivializarle, sobre todo si se tiene en mente el papel jugado por la idea de las “preferencias” atribuibles a su agente modélico: el maximizador autointeresado. Creo que dicha pérdida de opacidad de las preferencias, con la consiguiente apertura de éstas a la crítica, se debe a que, a diferencia de Frankfurt y, marcando un contraste todavía mayor tratándose de Gauthier, el agente tal y como lo piensa Taylor contaría, para llevar a cabo su labor de evaluación, reflexión y deliberación, con el *plus* de unos (nuevos) recursos de análisis y de crítica, los cuales serían completamente inaccesibles para el maximizador de Gauthier, e incluso no le serían del todo permitidos al agente-evaluador “solitario” de Frankfurt. Pues el caso es que con Taylor se piensa un horizonte valorativo que opera como el trasfondo indispensable de los juicios y decisiones de las personas, de manera que los sujetos que así juzgan y deciden, por una parte no pueden ser pensados *in abstracto*, o careciendo de vínculos con un tal horizonte y, por la otra, dichos vínculos se entrelazan con unos referentes normativos y de sentido, los cuales, a su vez, y por su misma naturaleza, son compartidos por el agente con *otros sujetos*, o son asumidos por éste como aquello que compartiría con esos otros.

Esto nos lleva a recalcar otro elemento de contraste para pensar lo que creo admite ser visto como una compleja fuente de recursos reflexivos y deliberativos de la que podría disponer el evaluador fuerte de Taylor, si se lo compara con aquellos recursos con los que contarían los tipos de agentes estudiados en los anteriores capítulos: a saber, ese horizonte valorativo con el que se relaciona el evaluador de Taylor, horizonte que dista de ser equiparable, tanto a la línea jerárquica de deseos y metadeseos que —como vimos en Frankfurt— terminaba por ser inexplicable, como a la casi vacuidad de recursos deliberativos a la que podría quedar condenado el agente de Gauthier. Si frente a estos dos autores el horizonte valorativo que nos presenta Taylor debe más bien ser pensado como jugando el papel del trasfondo supuesto por los juicios y elecciones del agente, pienso que entonces sería más acertado que se lo considerara como un complejo entramado de sentido, una suerte de “red” de naturaleza holista que, por definición, invita al análisis, al desvelamiento de aquellos referentes normativos a los que el agente apela al emitir sus juicios y, por tanto, también cuando evalúa sus

motivaciones y deseos. Creo que, aunque Taylor no lo diga expresamente, esta especial configuración del horizonte valorativo implicaría que en que dicha “red” de significatividad su evaluador fuerte hallaría los referentes de sentido y de valor que supone como válidos cuando juzga o apela a aquello que arriba ya se menciona como:

- 1) Lo que, en un sentido epistémico, *cree* que de hecho *ocurre*, o cabría esperar que ocurriese en el mundo, 2) Lo que piensa que *deberían*, en un sentido *moral*, hacer los agentes humanos, incluyéndose ella/él en dicho mundo, y en el ámbito de dichos deberes/exigencias, y 3) Lo que considera *valioso o importante* como parte de su idea de vida lograda, vida de la que, a diferencia de lo que piensa Frankfurt, no tendrían por qué quedar excluidas las vidas de otros agentes.¹⁶⁶ En relación con este asunto, repito, si bien Taylor parece hacer una marcada insistencia en el tercer tipo de referentes normativos, no obstante pienso que habría razones para incluir los otros y, además, para considerar que los diversos grupos de referentes tendrían de alguna forma que entrelazarse entre sí o, al menos, no se deberían establecer escisiones tan radicales entre ellos, dado precisamente que hablamos de un “horizonte”, de un “trasfondo” de naturaleza holista, de una “red” de sentido asociada a la vida del agente que, como veremos, es “una” vida.

Esta naturaleza holista del horizonte valorativo, a su vez relacionada con el tipo de lazos que con él tienen los agentes, explicaría por qué para éstos sus compromisos valorativos más importantes, a pesar de lo determinantes que resulten para sus juicios y elecciones, sin embargo no aparecen como completamente claros para los sujetos y de allí que los agentes requieran, sobre todo en algunas ocasiones acuciantes de su vida, llevar a cabo un proceso de *aclaración* de tales compromisos. Este proceso, que también es de autoaclaración en cuanto implica una indagación sobre la propia identidad, aunque Taylor no lo exprese en los siguientes términos (que acaso serían más propios de la ética del discurso), creo que necesariamente transcurre en el contexto de la comunicación, de un *diálogo* con *otros* agentes; con aquellos con quienes se pueden discutir las claves de normatividad y de sentido supuestas en ese trasfondo compartido. Éste, por tanto, insisto en que proveería una riqueza de recursos para la deliberación, la reflexión y

¹⁶⁶ Retomando lo advertido arriba, y como se comentará más adelante, estos tres órdenes normativos no tendrían por qué agotar aquello que el sujeto, si éste es pensado bajo el modelo de Taylor, supone en sus juicios y evaluaciones como “válido” o “valioso” o “importante”, ya que en ello también podría incluirse v.g. sus valoraciones acerca de lo bello.

discusión que, como arriba se menciona, pienso que no aparecería en el relato de Frankfurt pero, sobre todo, sería impensable dentro de una propuesta que, como la Gauthier, gira en torno de aquello a lo que me he referido como un modelo “reducido” del agente humano y de sus formas de deliberación. Sin embargo, es importante advertir que, al tratarse de un entramado holista, el proceso de la aclaración del mencionado horizonte no sería agotable ni estaría determinado del todo ni para siempre, puesto que este último sería susceptible de ser articulado, examinado y discutido una y otra vez, sin que se pueda señalar un término o un punto final para dicha aclaración, lo cual creo que marcaría una diferencia importante con lo que se estudió en el caso de Frankfurt. Adicionalmente, creo que lo anterior no debería hacer que le temiésemos a la falibilidad que, desde Taylor, y de nuevo atreviéndome a sugerir sus posibles lazos con posturas que le lleven más allá de su preocupación por el aspecto valorativo¹⁶⁷, pienso que habría que atribuirle a los agentes involucrados en el mencionado proceso de articulación, aclaración y discusión.

Otra advertencia importante que creo conviene hacer si se tiene en cuenta, tanto la naturaleza atribuible a este proceso, como aquella que caracterizaría los compromisos valorativos que en él se desvelen, así como la calidad de los vínculos que unen a los agentes con tales compromisos, es que en Taylor, y como intentaré que se muestre en el presente capítulo, por *una parte*, no sería correcto decir que los agentes simplemente “eligen” estos últimos, lo cual, como enseguida aclararé, pienso que permitiría apreciar una conexión importante con el *caring* de Frankfurt. Para el filósofo canadiense, el tipo de relación que guarda un agente con sus compromisos valorativos excluiría que a éstos se los pudiera pensar como el producto de una “elección”, es decir, como algo que el agente determina, decide o se inventa de manera radicalmente creativa y al margen de lo que su contexto pueda aportarle. Por el contrario, desde Taylor, dichos compromisos tendrían que ser vistos como un hecho con el que se tiene que contar, como el innegable punto de partida, dada la propia formación del agente y en virtud de su papel como hablante de una comunidad.¹⁶⁸ No obstante, como veremos, de *otra parte*, creo que para

¹⁶⁷ Al respecto conviene recordar lo señalado por Karl O. Apel acerca del tema del falibilismo y su relación con las teorías de la verdad. Cf. Apel (1991, pp. 37-145).

¹⁶⁸ Creo que Taylor estaría de acuerdo en que, con respecto a estas afirmaciones, se recuerde los legados de G. H. Mead (1972, Part III, *The Self*), y L. Wittgenstein (Trad. 2002); el primero, en lo referente a nuestra génesis como “personas” que se constituyen como tales en el contexto de una sociedad; y el

Taylor esto no significa que los compromisos valorativos de una persona tengan que ser vistos como una suerte de “destino” ineludible, de lastre inamovible e inexplicable que pese inexorablemente sobre el individuo ni, mucho menos, que dicho “peso” resulte inarticulable para el propio sujeto, ni para los otros agentes con los cuales quepa discutir en torno a tales compromisos; y esto último, en cambio, sí creo que puede significar una distancia con Frankfurt.

En cuanto al *primer* aspecto que acabo de señalar, pienso que en Taylor encontramos una insistencia, análoga a la de Frankfurt cuando se refería a la naturaleza de las preocupaciones de las personas. Pues el primero, repito, muestra que los compromisos valorativos de un agente no son objeto de una creación *ex nihilo* ni de una elección en el vacío. Empero, de *otra parte*, creo que, a diferencia de Frankfurt, habría en el filósofo canadiense una suerte de confianza en la razón, por lo menos, en algo parecido al *logon didonai*, en la posibilidad del dar y escuchar razones, lo cual, a su vez, sería coherente con una idea de los compromisos valorativos que, a diferencia del *caring* de Frankfurt, sí creo que admitiría la inclusión de los aspectos normativo y dialógico. En virtud de esto a lo que me refiero como una cierta confianza en la razón –obviamente *no* entendida en el sentido restringido en que es asumida por Gauthier–, pienso que Taylor concibe a los agentes humanos como capaces de desvelar, de sacar a la luz sus compromisos valorativos y someterlos a un examen crítico gracias a las armas del lenguaje, y de la articulación y explicitación que éste permitiría.¹⁶⁹ Se trataría, además, de un lenguaje que, como todos, no sería un lenguaje privado¹⁷⁰ y de un examen que, repito, tampoco puede darse sin la escucha de las voces de otros agentes, bien sea que éstos se hallen físicamente presentes, bien sea que sus voces sean internalizadas por quien intenta aclarar los referentes que comparta con ellos.

segundo, apuntando al asunto de la forja de nuestra competencia como hablantes, como seres que sólo pueden aprender a hablar gracias a otros hablantes.

¹⁶⁹ En este sentido, pienso que en Taylor se podrían ver solucionados los problemas que, como vimos en el primer capítulo, Brandom le señala a Gauthier, con respecto a su noción de “preferencia cualificada”.

¹⁷⁰ De nuevo, recuerdo la lección de Wittgenstein, esta vez más relacionada con lo que éste ve como la imposibilidad de los lenguajes privados. Cf. Wittgenstein (Trad. 2002, §§ 199-202).

1. Con Frankfurt pero más allá de él: la importancia del asunto valorativo

En su artículo “What is Human Agency?”, Taylor busca responder a la pregunta de Frankfurt acerca de qué es aquello que estamos dispuestos a atribuirle a una persona, a un ser humano responsable que puede ejercer esa forma de agencia a la que consideramos específicamente “humana”:

I would like to explore [in this paper] what is involved in the notion of a self, of a responsible human agent. What is it that we attribute to ourselves as human agents which we would not attribute to animals? (1985c, p. 15).¹⁷¹

Además de mostrarse de acuerdo con Frankfurt en que lo que parece ser propio de esa particular forma de agencia a la que estamos dispuestos a calificar como “nuestra”, “propiamente humana” o de “personas”, es una capacidad para *evaluar* nuestros deseos, Taylor manifiesta su entusiasmo por los alcances de este descubrimiento de Frankfurt. Concretamente, por su idea de que dicha capacidad nos revela un aspecto *reflexivo* y *autoevaluativo* sin el cual no podríamos pensarnos como el tipo de agentes que creemos ser. En este punto, sin embargo, pienso que habría que señalar una diferencia que, como se verá, trae consecuencias de gran calado en el discurso de Taylor: para éste, y por contra de lo que afirma Frankfurt, lo que resulta interesante de las voliciones de segundo orden a las que este último se refiere, es que son el producto de una capacidad *valorativa*, adjetivo que, según lo estudiado en el anterior capítulo, Frankfurt no estaría en absoluto dispuesto a utilizar. Esta diferencia, a su vez, se relaciona con otro posible punto de contraste entre este último y Taylor, a saber, que dicha capacidad evaluativa, al ser también autoevaluativa, según el autor canadiense significa que cabría afirmarse algo que en Frankfurt no siempre resulta ser visto del siguiente modo: se trata de una evaluación que, amén de alcanzar a nuestros deseos, se dirige también a nosotros mismos. Esto es, involucra de manera importante *juicios de valor* sobre lo que somos como personas, y sobre nuestra *manera de vivir*, o el modo de vida con el que asociamos dichos deseos.

¹⁷¹ Al igual que procedí en el capítulo anterior, en esta parte dedicada a Taylor las referencias a su textos se harán dentro del cuerpo mismo del presente capítulo, no a pie de página, anotando el año de publicación y la paginación correspondiente.

El *plus* que comporta este asunto valorativo explica que Taylor intente complementar y radicalizar el aporte de Frankfurt. Para ello, el primero propone que se introduzca una distinción adicional a aquella que el segundo ya había establecido con sus definiciones de deseos de primer orden, y deseos y voliciones segundo orden. La distinción adicional propuesta por Taylor se daría, ya no a nivel de los *deseos* en cuanto tales, sino en el ámbito de nuestra propia *capacidad para evaluar* tales deseos. Es decir, en el centro mismo de esa actividad que, según Taylor, y a diferencia de Frankfurt, es también valorativa y judicativa y que, para el primero, desplegamos cuando reflexionamos en torno a la calidad o “deseabilidad” de nuestras motivaciones y deseos. El filósofo canadiense plantea, por tanto, la necesidad de establecer una distinción que, según él, se daría a nivel de nuestras *formas de evaluar*, distinción que se daría entre lo que él llama “evaluaciones débiles” y aquellas otras evaluaciones a las que califica como “fuertes”. Esta diferencia lograría, al decir de Taylor, complementar y radicalizar la distinción propuesta por Frankfurt, de modo que se pueda propiamente sacar a la luz un elemento que el primero considera insoslayable cuando se intenta establecer qué es lo que nos caracteriza como agentes humanos, esto es, precisamente como ese particular tipo de agentes que son capaces de evaluar sus deseos. Dicho elemento que, repito, viene determinado por las *diversas formas de evaluación* que podemos hacer, a los ojos del autor canadiense aun no quedaría incluido en el discurso de Frankfurt cuando éste ofrece su respuesta a la pregunta sobre qué es aquello que parece ser privativo de una agencia específicamente humana si se la compara con otras formas de agencia, o qué es lo que estaríamos dispuestos a considerar como definitorio de un “yo-humano” (*self*).

1.1 Evaluaciones “débiles” vs. evaluaciones “fuertes”

Según Taylor, cuando hacemos evaluaciones *débiles* llevamos a cabo una labor deliberativa que nos permitiría decidirnó ante alguno de los siguientes *dos tipos de cuestiones* bastante similares entre sí: 1) Identificar, de entre dos o más deseos de los que tenemos, cuál es el que “nos mueve” con “más fuerza” o el que experimentamos como el “más urgente”; o también, si se prefiere, establecer cuál de entre dos o más objetos de elección o de deseo, es aquel que “nos atrae más”; y 2) Planificar estratégicamente un estado de cosas que sea lo más cercano posible a un óptimo de satisfacción de nuestros deseos, cuando se nos dificulte atender simultáneamente a dos o

más de ellos. Pues puede ocurrir que, por causa de las circunstancias del momento, no nos resulte fácil hacerlos compatibles entre sí y, por tanto, y a pesar nuestro, nos veríamos obligados a establecer a cuáles de esos deseos atender prioritariamente, y a qué otros sacrificar o aplazar. Tendríamos así que resignarnos con lograr un orden estratégico o un estado de cosas que resulte ser, a lo sumo, repito, el más cercano posible a ese óptimo de satisfacción que, idealmente, nos permitiría atender a todos los deseos en juego, pero que no siempre resulta ser un óptimo realizable.¹⁷²

Taylor propone, como ejemplos del *primer* tipo de deliberación en la que se hallarían involucradas evaluaciones débiles, aquella que llevaría a cabo una persona que está intentado aclararse a sí misma si le apetece más comerse un alimento del tipo X, o si preferiría más bien otro del tipo Y, v.g. uno dulce, o mejor otro salado. Algo bastante parecido le sucedería al agente que está buscando establecer si le atrae más pasar el próximo fin de semana en la playa o si, en cambio, preferiría irse a la montaña.¹⁷³ En cuanto al *segundo* tipo de asuntos que intentaríamos resolver cuando nos vemos obligados a hacer evaluaciones débiles, Taylor propone el ejemplo de una persona que, a pesar de que tiene hambre y le apetecería comer en estos momentos, está tratando de decidir si efectivamente le convendría hacerlo ahora, o si le vendría mejor posponer su almuerzo e irse primero a nadar, actividad que también le gustaría llevar a cabo. Aunque ambas opciones son atractivas para la persona, sin embargo ésta se ve ante la necesidad de identificar a cuál de las dos atender primero, pues sabe que si come ahora se le podría

¹⁷² Como puede verse, y aunque Taylor no lo diga utilizando estos términos, se buscaría gestionar una suerte de economía de los deseos, aplicando el principio de escasez, tal y como vimos que idealmente lo haría un agente pensado bajo el modelo “reducido” propuesto por Gauthier. Taylor no busca criticar a este último autor sino, como veremos, al utilitarismo, postura a la que Taylor acusa de excluir la posibilidad de una forma de agencia compleja como la que él propone. Empero, en lo que sigue puede verse que su descripción de las evaluaciones débiles, y su crítica al utilitarismo, al que Taylor ataca como la posición desde la que se asume dicha forma de evaluación como la única plausible, arrojan una luz interesante y de gran utilidad para ver más de cerca los fallos del modelo restringido de Gauthier. Dicho modelo, aun en contra de lo que este último sostiene, como vimos en los primeros capítulos, pienso que resulta bastante cercano al modelo utilitarista criticado por Taylor, crítica sobre la que nos ocuparemos enseguida. No en vano, recordemos, Gauthier defiende la tesis de que las decisiones morales, si han de ser pensadas como racionales, deben reducirse a un cálculo económico, lo cual, como se comentará, es duramente descalificado por Taylor en su análisis del modelo utilitarista de agencia humana.

¹⁷³ Ello suponiendo, claro está, y como veremos a continuación, que en estas disyuntivas *no* se hallen implicadas cuestiones a las que el agente otorgue un *especial significado* para el desarrollo de *su vida*. Por ejemplo, que ir a la montaña consista acaso en llevar a cabo una suerte de rito de purificación ordenado por un compromiso religioso al que el agente considera de la mayor importancia; o que comerse algo dulce implique un peligroso atentado en contra de su salud, dando por sentado, de nuevo, que la persona igualmente se preocupe por ésta.

hacer difícil nadar luego, pero también le ocurre que siente hambre y, a lo mejor, el horario en el que estará abierta la piscina podría permitirle comer ahora y nadar después. Así, el sujeto tendría entonces que establecer cuál de los dos cursos de acción llevar a cabo primero y cuál debería venir posteriormente, dependiendo de factores empíricos tales como el horario en el que va a estar abierta la piscina, la distancia hasta ella, la urgencia que sienta la persona por comer y el tiempo que le tomaría recuperarse de un copioso almuerzo si luego se fuera a nadar (1985c, pp. 16 y ss.).¹⁷⁴

Lo que, según Taylor, ilustrarían estos ejemplos, tanto los de la primera categoría, como los de la segunda, es la característica común a *todas* las evaluaciones *débiles*: en ellas la deliberación del agente *no* involucra juicios de *valor* sobre las alternativas a elegir. Ninguna de las opciones que se le presentan a la persona es vista por ella como poseyendo un valor especial si se la compara con las otras. De manera que su elección se lograría examinando solamente factores tales como lo empíricamente alcanzable y el grado de atracción que ejerzan las distintas posibilidades que se le ofrecen. En los ejemplos mencionados, a menos que irrumpa un asunto de vital importancia para la persona, v.g. una prescripción médica que desaconseje la ingestión de alimentos salados, o una promesa que el sujeto quiere cumplir, en principio éste no tendría por qué pensar que haya algo especialmente valioso en comerse un alimento dulce si se lo compara con tomar otro salado, o en irse al mar, en comparación con lo que sería para la persona pasar este fin de semana en la montaña. Para Taylor, cuando se está ante este tipo de circunstancias en las que la decisión depende del planteamiento de las opciones en términos de evaluaciones *débiles*, *no* habría en juego un *significado especial* para el agente, de forma que éste *no* tendría por qué considerar que *su vida*, o que *él mismo como persona*, sean “mejores” ni “peores” si escoge alguna de las alternativas que se ofrecen ante él.¹⁷⁵

¹⁷⁴ Estoy haciendo una paráfrasis de la forma en que Taylor expone los ejemplos, y éstos, a su vez, no son exactamente los mismos que él propone, pero creo que sería aconsejable hacerlo para superar un poco la barrera del idioma y para capturar lo que pienso que el autor intenta decir (Cf. 1985c, pp. 16 y ss.).

¹⁷⁵ De nuevo, es posible que alguien quisiera llamarle la atención a Taylor sobre un eventual prejuicio suyo, manifestado en estos ejemplos que él propone. A saber, que nuestro autor parecería pasar por alto que bien puede haber contextos en los que cuestiones que parecen banales para una persona, puedan no serlo para otra. Por ejemplo, el uso de determinadas prendas de vestir en ciertos espacios y en ciertas culturas podría tener connotaciones importantes que, en cambio, no se darían en el contexto de otras culturas y tratándose de otras personas. Sin embargo, creo que, hasta aquí, podría obviarse “la letra” de los ejemplos propuestos por Taylor y seguir avanzando en su discurso, concediéndole, por ahora, “el fondo” o “el espíritu” de la distinción que intenta presentarnos entre evaluaciones “débiles” y

Todo lo contrario ocurriría, según el autor, en aquellos casos en los que nos vemos en la situación de tener que hacer evaluaciones *fuertes*. En ocasiones de esta naturaleza, y en franco contraste con los ejemplos anteriores, nuestra deliberación sí involucraría juicios de *valor* o, mejor, dicha deliberación tendría por objeto intentar establecer el valor que tienen nuestros deseos o que tendrían las opciones entre las cuales debemos elegir. El esquema propuesto por Taylor es el que normalmente seguiría la deliberación y la decisión de una persona que se ve ante la necesidad de elegir entre actuar de una cierta manera X, o de otra manera no-X pero, en este caso, la diferencia importante frente a los ejemplos anteriores está en que la deliberación del individuo sí involucra un juicio sobre la calidad y el valor que la persona le atribuye, no sólo a las acciones entre las que tendría que elegir, sino también a las *motivaciones* que estarían tras los cursos de acción por los que debe optar. Así, por ejemplo, la persona podría caer en la cuenta de que si hiciera X, en realidad lo haría porque a ello le empujaría un cierto deseo al que juzga como poseyendo una cierta naturaleza valorativamente relevante, y que a los ojos de la persona resulta oponerse a ese otro deseo que le llevaría, por el contrario a realizar la acción no-X. De este modo, el agente puede, finalmente, decidirse por no hacer X porque se ha percatado de que si lo hiciera, lo haría llevado por *un determinado tipo de deseo* al que *valora* negativamente apelando a calificativos *fuertes*, tales como “malo”, “dañino”, “innoble”, “bajo”, dado que se trataría v.g. de un sentimiento de envidia, o de un deseo de venganza. Este deseo es valorado por el agente como radicalmente opuesto al deseo que le empujaría a llevar a cabo la acción contraria a X, es decir, un deseo que es juzgado por el agente mediante el uso de términos – también valorativamente cargados o fuertes– tales como “noble”, “decente”, “correcto”, que serían los apelativos opuestos a aquellos a los que acude cuando juzga el deseo que le atraería hacia la acción X.

evaluaciones “fuertes”, sobre las que enseguida se hablará. Pues creo que el autor busca decirnos que la diferencia está en *el hacer* esta clase de *evaluaciones*, independientemente de los *objetos concretos* de elección o de los deseos a los que se evalúe, bien sea de manera *débil*, bien de manera *fuerte*. Pienso que, para Taylor, *hacer ambas clases de evaluaciones*, y *experimentar la diferencia* entre ellas, sí sería común a diversos sujetos y culturas, por más que, insisto, en cada una de estas culturas y tratándose de distintas personas, cambien los objetos específicos de juicio. Así, lo que interesaría mostrar, por parte del autor, es que, al margen de la cuestión de hecho de que se trate de montaña/playa, dulce/salado, nadar/comer, el caso es que *hacemos la distinción* entre unas elecciones que comportan un *valor* importante para nosotros, y otras elecciones en las que no hay tal valor en juego, como podrá verse en lo que sigue de su exposición.

Para ilustrar este esquema, uno de los ejemplos propuesto por Taylor es el de una persona que se ve ante la disyuntiva de huir ante un peligro o de enfrentarse a éste, cuando es el caso que el agente vería lo primero como un acto de *cobardía* y, por el contrario, juzgaría a la segunda opción como un acto de *coraje/valor*. Así, no sólo son objeto de un juicio valorativo las *acciones* mismas, o lo que con ellas pueda obtener la persona, o el *resultado* de tales acciones, sino que también son valorados de este modo el deseo o la *motivación* que la persona asocia a tales actos (y sus resultados). De esta manera, el agente no sólo leería la situación en términos tales como “salvar mi vida”, “salvar mi empleo”, “dar una apariencia favorable ante determinadas personas”, o por el contrario, “arriesgar mi vida”, “salvar mi honor”, “salvar mi patria”, “sentirme digno/orgulloso de mí mismo”, “no aparecer como un cobarde/un adulator ante ciertas personas”, etc. Todos estos *resultados*, así como las *acciones* de las que empíricamente podrían derivarse, son valorados de manera “fuerte” por el sujeto porque igualmente éste valora las *motivaciones* que le llevarían a elegir alguno de esos cursos de acción (o alguno de esos resultados esperables) entre los que se ve obligado a elegir, esto es, en el caso del ejemplo propuesto por Taylor, enfrentarse a un peligro o huir ante él. Así, nuestro autor cree que si la persona lee estas acciones de manera fuerte es porque también ve en ellas la expresión de un cierto valor y del correspondiente antivalor, en este caso, el coraje/la valentía *vs.* la cobardía. De allí que la persona utilice, para describir las acciones que podría realizar y entre las que tendría que decidirse, términos que indican una “mayor” o “menor valía”, una jerarquización que separaría lo “más alto” de lo “más bajo”. Por ejemplo, diría del acto X que es el más “noble”, “decente” o “meritorio” y, por el contrario, del acto contrario no-X, que es la acción más “baja”, “indecente”, “ruin” o “vergonzosa”.

Aunque Taylor no lo explicita, creo que sí se encuentra advertido de la siguiente objeción: puede ocurrir que haya personas que, o bien se decidan por aquello que consideran francamente más “bajo” y no les importe hacerlo; o lo hagan aunque se sientan mal por ello; o lo hagan y, por el contrario, se sientan perfectamente disculpadas por haberlo hecho; o bien podría tratarse de individuos a los que sencillamente les tiene sin cuidado qué sea “bajo” o qué no lo sea; o incluso podría darse el caso de personas para las cuales ni siquiera tendrían sentido a este tipo de calificaciones valorativamente cargadas. Si bien, repito, Taylor no expone esta objeción, creo que la tiene presente y

ante ella él insistiría que, aún siendo posibles los hechos que se toman como base para cuestionar su idea del uso de un vocabulario valorativamente cargado, de todas formas es igualmente posible –y tenemos experiencia de ello– que se presenten circunstancias de elección en las que a ciertos agentes sí le resulten significativas estas expresiones valorativas fuertes, y en las que éstos le asignen un valor a las opciones a elegir. Así, si es el caso que la persona no considera para nada irrelevante el *valor* que le atribuye a dichas opciones, estaríamos, según nuestro autor, ante unas circunstancias en las que el agente lleva a cabo lo que Taylor llama “evaluaciones fuertes”. Ahora bien, aunque el filósofo canadiense en ocasiones no sea muy enfático en afirmarlo, igualmente creo que estaría de acuerdo en que este tipo de evaluaciones *no* tendrían por qué reducirse al ámbito en el que los agentes hacen valoraciones de tipo *moral* y de tipo *ético*. Pienso por tanto que Taylor incluiría, dentro de las diversas formas en la que podríamos hacer evaluaciones fuertes, también aquellas que atañen a *otros* órdenes normativos. Por ejemplo, aquellos que se ejemplificarían en nuestros juicios sobre “lo elegante” como opuesto a “lo vulgar”, “lo original” como opuesto a “lo trillado”, “lo sencillo” como opuesto a “lo artificioso/intrincando”, teniendo en cuenta el vínculo entre estas expresiones y la actitud de la persona que las utiliza, así como su lectura del contexto en el cual lleva a cabo tales juicios valorativos.

1.2 La idea de “forma de vida”

Creo que para Taylor, trátase o no de un valor específicamente ético, lo importante es que cuando el agente juzga la situación de tal manera que cree que en ella estaría en juego un valor ante el cual no es indiferente, entonces el sujeto se vería a sí mismo como afectado por dicho valor, en el sentido de que juzgaría que la expresión o la negación de tal valor serían *reveladoras del tipo de persona que es*, o que quisiera ser, así como del *modo de vivir* que quisiera fuera el suyo.¹⁷⁶ Este es, precisamente, otro de los aspectos que distinguiría a las evaluaciones fuertes de las débiles: a saber, que las motivaciones y los deseos que la persona somete a juicio cuando lleva a cabo el primer tipo de evaluaciones, son valorados “fuertemente” porque el sujeto los asocia de una

¹⁷⁶ Pienso que esto, igualmente, al menos en parte, y aunque Frankfurt no estuviera de acuerdo con el uso del término “valor”, tendría que ser visto como aquello que daría sentido a su idea de que somos seres para los cuales las cosas *importan*, a los que les *preocupan* ciertos asuntos.

manera *no accidental* con ciertas *formas de vida*. Dichos deseos y motivaciones serían vistos por el agente como haciendo parte de tales formas de vida, o como siendo especialmente expresivas de éstas. Luego, dichas formas de vida, al igual que las motivaciones y los deseos con las cuales el agente las asocia de manera no accidental, son juzgadas por éste como modos de ser persona y como maneras de llevar una vida a las que también calificaría utilizando un lenguaje valorativamente cargado. Se trataría de un vocabulario fuerte como el que se ejemplifica en expresiones tales como “una vida frívola”, o “santa”, o “desvergonzada”. Con el uso de este tipo de expresiones el agente manifestaría un juicio sobre las maneras de vivir a las que admira, o en las que quisiera tomar parte, o con las que se identifica, al tiempo que establecería una oposición entre éstas y esas otras maneras de vivir a las que, por el contrario, la persona considera poco admirables, de las que quisiera apartarse, o en las que no quisiera participar mediante sus acciones y elecciones y, por tanto, a las que asocia con deseos a los que igualmente evalúa negativa y fuertemente como desposeídos de cierto valor, o como expresiones de cierto antivalor.

Taylor resalta que el nexo que el agente establece entre un deseo al que evalúa negativamente y la forma de vida a la que la persona asocia dicho deseo, a diferencia de lo que ocurre con los deseos y opciones de elección sobre los que delibera quien está llevando a cabo una evaluación *débil*, *no* es un nexo *accidental*. Si el agente no pudo ir a nadar ahora porque tiene hambre y prefiere comer primero, es sólo porque casualmente sucede que ahora se siente hambriento, pero bien podría haber ocurrido que no se sintiera de ese modo y, por tanto, en ese caso podría haberse ido a nadar. Pero algo muy distinto ocurre tratándose de una persona que toma la determinación de quedarse a enfrentar un peligro *porque* ve en esa acción la *expresión de un cierto valor*. En su caso, decide no huir *no* porque en estos momentos no pueda hacerlo a causa de una circunstancia empíricamente superable, por ejemplo, que su autobús tardará en llegar. Si decide permanecer en su sitio y no huir ante el peligro, es porque el acto mismo de huir es visto por la persona como algo que considera como *esencialmente* incompatible con una cierta *manera de vivir* en la cual quisiera participar, y quisiera hacerlo justamente porque considera que en ella se privilegia o se expresa un cierto valor, en este caso, el del coraje/la valentía, valor que no es accidentalmente opuesto a la cobardía que la persona ve expresada en el acto de huir.

Por lo anterior, otra de las diferencias importantes que, según Taylor, separaría a las evaluaciones débiles de las fuertes, viene dada por el hecho de que cuando se evalúan débilmente las opciones a elegir, el agente puede considerar como “buena” alguna de ellas simplemente por el hecho de que la desea, mientras que si se encuentra en una situación en la cual evalúa fuertemente sus alternativas, para que considere a alguna de ellas como “buena” no basta simplemente con el hecho de que la desee; incluso puede llegar a suceder que no es propiamente uno de sus objetos de deseo y, a pesar de ello, dicha alternativa de acción es juzgada por la persona como “buena”. Dicho juicio, para nuestro autor, ancla en la relación no accidental que, a los ojos del agente, une a la opción así valorada fuertemente (como buena, o como “la” buena”) con un determinado modo de vida, al que, como acabamos de ver, la persona evalúa fuertemente en el sentido descrito. Ahora bien, para llevar a cabo dicha evaluación, el agente necesita hacer uso de un lenguaje valorativamente cargado, que le permita *articular* sus opciones de elección en términos como los ya comentados, esto es, como los que ejemplifican denominaciones “fuertes”, tales como “cobarde” y “valiente”. Esto nos lleva a otro elemento muy importante que marcaría un contraste de gran significado entre evaluar fuertemente y hacerlo, por contra, débilmente: la distancia que habría entre el *lenguaje* que despliegan los agentes cuando hacen el primer tipo de evaluación y el lenguaje del que hacen uso cuando evalúan en un sentido débil. Sobre este tema, que resulta, como digo, de gran importancia, nos ocuparemos en breve. Antes de continuar con él, en este punto me atrevo a abrir un breve paréntesis que creo puede ser útil para apreciar la importancia que tiene la idea de Taylor de que las evaluaciones fuertes se dirigen también a *modos de vivir*, que el agente asocia de una manera no accidental con ciertas motivaciones y ciertos actos.

Expresiones tales como “modos de vivir” o “formas de vida” creo que serían indicativas de *referentes complejos* cuya naturaleza acaso estaría mejor capturada por la imagen de una red, que por la de una línea jerárquica, tal y como esta última aparece en el relato de Frankfurt. Aun concediendo que Taylor igualmente menciona oposiciones como la que habría entre “lo más alto” y lo más “bajo”, y aceptando que valorar fuertemente incluya también establecer ciertas preferencias (a las que podríamos llamar “fuertes”)¹⁷⁷ y, por tanto, ciertas relaciones que admiten ser vistas como “de

¹⁷⁷ Habermas (1990) utiliza la expresión “preferencias fuertes” inspirándose precisamente en Taylor.

mayor/menor prelación” o de jerarquía, en síntesis, aceptando que hay un lugar válido e importante para este tipo de relaciones que expresan una jerarquía, sin embargo pienso que habría que evitar aferrarse sin matices a la imagen de una línea jerárquica, y más bien aprovechar el recurso metafórico de la red. Pues este último, a diferencia de la primera, sí capturaría la idea que, a mi entender, es la importante y es la que permitiría apreciar la necesidad de señalar que somos seres que evalúan sus deseos, así como comprender por qué lo hacemos, y por qué algunas de esas evaluaciones implican *juicios de valor*: a saber, la noción de *formas de vida*.

Desde esta noción creo que podría hacérsenos más fácil comprender, no sólo el propio esquema jerárquico o el establecimiento de las mencionadas jerarquías, sino también la naturaleza y el sentido de la evaluación de los deseos, así como la importancia que reviste el aporte del *caring* propuesto por Frankfurt, en tanto que elemento fundamental de nuestro particular modo de agencia. Y, por el contrario, pienso que esta comprensión se nos haría más difícil si nos apegamos a la imagen de la línea jerárquica, imagen que, recordemos, en buena parte era la causante de que el mecanismo (palabra que ya traería sus dificultades)¹⁷⁸ de la evaluación de los deseos fuese objeto de serias sospechas, tales como las que se manifiestan en las críticas de Watson a Frankfurt. Por contra, creo que la idea de una red compleja de referentes normativos sería más adecuada para que, desde ella, el agente pueda dar razón de eso que, en términos de Taylor, serían sus “evaluaciones” –desde las menos hasta las más importantes– y por tanto, también de aquello que Frankfurt vería como sus “preocupaciones” fundamentales, las cuales, si retomamos críticas como las que Bratman y Scanlon le hacía a este último, aparecerían igualmente como necesitadas de explicación.¹⁷⁹

Otro panorama menos halagüeño se abre ante nosotros si insistimos en la imagen de una jerarquía lineal, ya que terminaríamos por llegar al mismo resultado señalado por Watson, esto es: las voliciones que están “más arriba” (en esa jerarquía) serían inexplicables al no haber manera de conectarlas con otros referentes con los que se entrelacen y que permitan dar cuenta de ellas. De manera similar, el agente no sería

¹⁷⁸ Remito al lector a las críticas, comentadas en el capítulo primero, que Brandom y Ripstein le hacen a Gauthier, en cuanto al aspecto reduccionista de su discurso.

¹⁷⁹ A menos, claro, que decidamos atrincherarnos en una inefable “necesidad volitiva”.

capaz de explicar por qué se preocupa por aquello que se preocupa. En síntesis, no podríamos hacer inteligible por qué evaluamos nuestros deseos, lo cual, amén de hacernos sospechar de la línea jerárquica como imagen, pienso que también nos invita a sospechar del *tipo de justificación* que puede ofrecer quien se tome demasiado en serio dicha imagen. Me refiero a que podríamos caer en cuenta de que para ofrecer una justificación de por qué consideramos como “indeseables” determinados actos, o el tipo de deseos que nos moverían a realizarlos, no tendría mucho sentido apuntar a un metadeseo de orden superior, como si se tratara de llevar a cabo una suerte de demostración apodíctico-deductiva, en la que para validar una proposición se apele a un cierto axioma de orden “superior” del que se supone “se deduce” la validez de la proposición que se está examinando, axioma que, a su vez, no sería susceptible de una demostración *sensu stricto*.¹⁸⁰

Por lo anterior, creo que el sentido, la razón de ser de ese juicio valorativo se nos haría más accesible si, en *primer* lugar, y a diferencia de lo dicho por Frankfurt, dicho juicio se reconoce como tal *juicio*, y como “valorativo” y, por tanto, como implicando la apelación a referentes normativos. En *segundo* término, un juicio valorativo mediante el cual un deseo/acto es considerado como deseable/indeseable se haría más comprensible si dicho deseo/acto es contextualizado mostrándose como imbricado con otros deseos/actos con los que, a su vez, conformaría una compleja red de referentes que se solicitan mutuamente, y que constituirían aquello que, en mi opinión, sería visto por Taylor como eso a lo que se refiere como “modo de vida”. Si bien es cierto que aquí Frankfurt podría señalar su significativo aporte del *caring*, de las “preocupaciones fundamentales”, insisto en que éstas no aparecen en su discurso como haciendo parte de un marco referencial complejo en el que cobren sentido, se relacionen entre sí, o pueda, en general, darse razón de ellas. De allí que, a mi entender, se requeriría del *plus* de la cuestión del valor y de las formas de vida que hallamos en Taylor.

¹⁸⁰ Al respecto, resulta muy sugerente la crítica que hace el propio Taylor (1985 y 1997a) en contra de los intentos por implantar este modelo deductivo en las ciencias sociales.

1.3 El lenguaje de las evaluaciones fuertes

El juicio que emite un agente sobre la deseabilidad/indeseabilidad de cierto deseo/acto, como se ha anunciado, se relaciona con otro asunto que reviste gran importancia para el filósofo canadiense: las *formas de lenguaje* que son propias de los diversos modos de evaluar los deseos y los modos de vida a los que éstos se asocian. Este tema creo que tampoco tendría sentido en el contexto del discurso de Frankfurt en tanto que éste, desde su postura anticognitivist, persista en su empeño de expulsar del *caring* la cuestión *normativa* y, en consecuencia, en seguir atado a su modelo *monológico* de agente. Este asunto nos lleva a una observación, a mi entender, importante, hecha por Taylor en su intento por describir las diferencias asociadas a cada una de las formas de evaluación, las evaluaciones fuertes y las débiles. El autor canadiense afirma que el tipo de *lenguaje* que debe desplegar el agente cuando lleva a cabo el primer modo de evaluar, al tratarse de un lenguaje en el que se expresan juicios de valor, juicios en los que la realidad es vista como conformada por cualidades diversas e irreducibles entre sí, sería un lenguaje de mayor riqueza, lo cual, a su vez, le permite al agente acudir a un repertorio más amplio y variopinto mediante el que alcanza a *articular* tales juicios, a dar cuenta de los diversos elementos que toma en consideración cuando delibera sobre esos asuntos a los que valora “fuertemente”.

De este modo, si el agente tuviese que *explicarnos por qué* no huyó cuando tuvo la oportunidad de hacerlo, tal vez pueda *contarnos más cosas*, relacionadas, repito, con esa *manera de vivir* que quisiera fuera la suya. Mientras que, tratándose de una evaluación débil, tal como aquella que se ejemplifica en casos como el de la persona que está tratando de decidir si ir a nadar o comer primero, si merendar algo dulce o algo salado, según Taylor, el tipo de lenguaje que el agente necesita para dar cuenta de su deliberación y de su elección final, no requiere de suyo la riqueza de recursos expresivos que, para nuestro autor, se evidencia en el lenguaje propio de las evaluaciones fuertes. De este modo, el intento por ofrecer una razón relacionada con una evaluación débil, termina, según Taylor, en la experiencia de la “inarticulación”, en la dificultad para explicitar motivos que, al no tocar cuestiones valorativas fuertes, al no involucrar el compromiso con una forma de vida, con una red compleja de sentido, terminan por ser “razones” del tipo: “porque sí”, “porque así me apetece”.

Esto último, como puede verse, creo que nos llevaría de nuevo a Frankfurt y al problema que sus críticos le señalan: la *imposibilidad de dar razones* para las voliciones de orden superior y para las preocupaciones fundamentales, imposibilidad que iría en contravía del carácter reflexivo de su agente modélico y del importante papel que el propio Frankfurt le atribuye a las preocupaciones fundamentales entendidas como “norte vital” de las personas. Pienso, por tanto, que dicha imposibilidad estaría atada a esta *inarticulación* que, para Taylor, caracterizaría las expresiones mediante las que se da cuenta de las evaluaciones débiles. Ambos aspectos: 1) *Referentes complejos*, tales como las formas de vida, y 2) *Formas de lenguaje más ricas*, estarían relacionados entre sí. Y creo que dicha relación nos ayudaría a entender un poco mejor el escollo que supone ese *mutismo* en el que parece que Frankfurt quisiera instalar a su agente modélico si éste fuera cuestionado sobre sus razones para actuar. Sobre este tema volveremos más adelante, cuando intente poner bajo la luz de Taylor los ejemplos mediante los cuales Frankfurt busca explicar su noción de “lo impensable”, noción que, a mi parecer, se relaciona con eso a lo que aquí me permito llamar “mutismo”. Pero antes de llegar allá, tendré que detenerme un poco más en el asunto del *lenguaje* asociado, tanto a las *evaluaciones fuertes*, como al *tipo de agente* que, según Taylor, llevaría a cabo tal forma de evaluación.

1.4 El lenguaje del evaluador fuerte vs. el del maximizador. Lo cualitativo y lo cuantitativo

Pienso que Taylor expone su idea de las evaluaciones fuertes al filo de una importante discusión que plantea en contra del *utilitarismo*, a causa del modelo de *agencia* que, según el canadiense, se privilegia desde dicha postura. Según nuestro autor, este modelo está estrechamente relacionado con una cierta idea de la *deliberación moral* y, de manera fundamental, con el tipo de *lenguaje* que debería utilizar el sujeto que lleva a cabo dicha deliberación. De este modo, creo que la tesis de Taylor de que existe una diferencia importante entre hacer evaluaciones débiles y hacer evaluaciones fuertes, en buena parte se entiende como un reclamo en contra del tipo de *agente* modélico que le atribuye a discursos como el utilitarista, al que nuestro autor considera representativo de una forma bastante exitosa, pero también incorrecta, de entendernos a nosotros mismos como agentes humanos. Pienso que dicho éxito y dicho error para

Taylor serían de tal magnitud que incluso el autor vería en el utilitarismo lo que, en general, acaso constituya uno de los motivos de análisis y de preocupación más importantes de su trayectoria filosófica: la idea de que nosotros, como occidentales modernos, tenemos acerca de nosotros mismos en tanto que agentes humanos o, mejor, ese ideal de agente al que, como miembros de nuestra cultura y de nuestros tiempos, tal vez quisiéramos llegar a parecernos.¹⁸¹

Dicho ideal, en tanto que problema filosófico, ha sido objeto de diversos desarrollos por parte de Taylor y, como digo, ha constituido una preocupación muy importante para él, incluso hasta sus obras más recientes. De modo que creo que, como arriba lo menciono, la lectura que pueda hacerse de su tesis acerca de las evaluaciones fuertes y, como veremos, acerca del tipo de agente que llevaría a cabo tales evaluaciones, lectura que en principio uno podría delimitar al artículo que aquí se está comentando, “What is Human Agency?”, tal vez se aprecie mejor teniendo como trasfondo su crítica, tanto al utilitarismo como, en general, a una amplia gama de posturas filosóficas que, para Taylor, participan del espíritu reduccionista y naturalista que según él motiva al utilitarismo como corriente. Dicho espíritu tendría en el modelo de agencia humana uno de sus productos más importantes y, a los ojos de Taylor, creo que también uno de los más preocupantes, precisamente porque no se trata de un espíritu que simplemente se atrinchere en las corrientes más marginales (menos exitosas, menos conocidas, menos populares) del pensamiento occidental sino que, por el contrario, ha hecho parte de movimientos bastante influyentes en nuestro desarrollo cultural, y no sólo en lo que se refiere a la filosofía y al mundo académico, sino también a las ciencias humanas, a las ciencias sociales y a la cultura popular. De manera que la polémica en contra del utilitarismo y su modelo de agencia humana creo que puede leerse como debida al hecho de que Taylor ve en dicha escuela un exponente claramente representativo de una cierta forma de vernos a nosotros mismos, forma con la que nuestro autor asocia buena parte de las sinsalidas y desaciertos con los que puede

¹⁸¹ Si bien Taylor (1994, 1996, 2006 y 2007) no comparte este ideal, al que el canadiense no sólo considera de naturaleza epistemológica, sino también como un ideal cuyas raíces son para él de carácter fundamentalmente moral, el autor intenta hallar la razón de ser de dicho ideal: el compromiso con una idea, indudablemente atractiva y en buena parte justificada desde el punto de vista moral, de una libertad, de una emancipación que nos conferiría, como seres humanos, una mayor dignidad asociada al señorío que tendríamos frente al mundo físico y al mundo social; incluso frente a aquello que, dentro de nosotros, parece atarnos a la naturaleza y las fuerzas psicológicas con las que a veces nos cuesta tanto lidiar.

encontrarse el pensamiento occidental moderno. Tales sinsalidas, según Taylor, tendrían su raíz en el carácter *naturalista* de teorías que, como el utilitarismo, comparten un ideal *reduccionista*, tal como el que vimos ejemplificado en Gauthier.¹⁸²

I wanted to argue against the understanding of human life and action implicit in an influential family of theories in the sciences of man. The common feature of this family is the ambition to model the study of man on the natural sciences. [. . .] What is striking about these family of theories is their reductive nature [. . .] this family of theories shares an allegiance to “naturalism”, by which I mean not just the view that man can be seen as part of nature –in one sense or other this would be surely be accepted by everyone– but that the nature of which he is a part is to be understood according to the canons which emerged in the seventeenth-century revolution in natural science. [. . .] Behind and supporting the impetus to naturalism . . . the understandable prestige of the natural science model, stands an attachment to a certain picture of the agent . . . it shows us as capable of achieving a kind of disengagement from our world by objectifying it (1985d, pp. 1-4).¹⁸³

En el contexto de “What is...?”, nuestro autor introduce su crítica al utilitarismo a propósito de la importancia que reviste para Taylor el tema de las evaluaciones fuertes y del lenguaje valorativamente cargado en el que dichas evaluaciones se expresan. Pues justamente el filósofo canadiense considera que desde una postura como la utilitaria tales tipos de evaluación y de lenguaje quedarían excluidos por completo. Para el autor, el utilitarismo presenta el problema de que postula como modélico un tipo de *agente* humano que sería *ajeno a lo valorativo* y, como en breve veremos, también a lo *cualitativo* en general, como opuesto al aspecto cuantitativo de los juicios y deliberaciones que haría dicho agente. Según Taylor, desde una postura como la utilitaria las evaluaciones fuertes no sólo no serían plausibles, sino que además serían un auténtico escollo que impediría la deliberación de un agente racional que intente tomar una decisión que, bajo criterios utilitarios, sea moralmente correcta. En otros términos, nuestro autor señala que si se asume una posición utilitarista, entonces tendría que suscribirse la tesis, a su parecer incorrecta y de importantes consecuencias, de que las únicas evaluaciones que podría hacer un agente que esté intentando adelantar un

¹⁸² Recordemos las críticas que, en este mismo sentido, recibe Gauthier por parte de Brandom y de Ripstein.

¹⁸³ Creo que estas palabras de Taylor nos recordarían la preocupación, ya señalada en el capítulo anterior, que autores como Strawson (1995) y Habermas (1989 y 1994a) manifiestan frente esta postura “objetivadora” –opuesta a la actitud participante– ante el mundo moral.

proceso de deliberación serían, en términos de Taylor, “débiles”, o sea, ajenas a la cuestión del valor.

Incluso, dicha cuestión sería proscrita del discurso utilitario ya que impediría que su agente modélico tome sus decisiones de tal modo que éstas obedezcan a un correcto cálculo de utilidad esperada. Recordemos que este tipo de agente debe ser pensado como un maximizador altruista¹⁸⁴ que busca obtener la mayor utilidad posible para la mayoría de los afectados por una decisión suya (incluyéndose a él mismo dentro de los así afectados). Para ello, debe llevar a cabo su proceso de deliberación de manera que, a partir del acopio necesario de información, esté en capacidad de describir sus opciones de elección de un modo tal que pueda “traducir” a una unidad común los costes y beneficios que quepa esperarse de cada opción. Sin esto, no podría identificar la alternativa de la que cabría esperarse una sumatoria de mayor utilidad para los afectados por ella. Se trata, pues, de una operación de cálculo que necesariamente implica borrar toda diferencia cualitativa y valorativa, con el fin de que dicho cálculo sea posible. Por lo tanto el maximizador de bienestar, tal y como aparece en el modelo utilitario, no debe en absoluto caracterizar sus opciones de elección acudiendo al uso de un *lenguaje cualitativo*, o un *lenguaje valorativamente cargado*, pues éste tornaría en inmensurables los “bienes” y los “males” –que en el argot de un maximizador deben ser expresados como “costes” y “beneficios”– que comporte cada una de sus opciones de elección. Criticando este modelo de deliberación que, repito, sólo puede darse a partir de lo que Taylor llama evaluaciones débiles, nuestro autor tratará de mostrar que si, bajo dicho modelo de deliberación las decisiones humanas serían reducidas a formas de cálculo, para lo cual es indispensable un lenguaje desprovisto de términos valorativos fuertes, entonces el modelo criticado sería indesligable del tipo de *agente* que haría uso de esa forma específica de lenguaje.

¹⁸⁴ Tomo esta expresión de un utilitarista contemporáneo: J. Harsanyi (1982, 1983 y 1985). Por supuesto, el maximizador *altruista* de Harsanyi no sería el mismo tipo de maximizador que propone Gauthier, quien postula, recordemos, como modélica, la figura de una agente *autointeresado* y critica, por contra, la idea de un maximizador de la utilidad social, tal y como vimos en los capítulos iniciales, sobre todo, en su crítica a Harsanyi. Sin embargo, creo que *ambos* maximizadores serían considerados modelos implausibles, harían parte del mismo error a los ojos de Taylor, quien arremete en contra del utilitarismo por la misma razón que pienso que bien podría hacerlo en contra de Gauthier: por su *modelo de agencia* humana.

The bent of utilitarianism has been to do away with qualitative distinctions of worth on the grounds that they represent confused perceptions of the real bases of our preferences which are quantitative. The hope has been that once we have done away with strong evaluation we will be able to calculate (1985c, p. 17).

Pienso que a los ojos de Taylor estarían, entonces, imbricados entre sí los modelos del *agente* (de *self*), de *deliberación práctica* y de uso del *lenguaje* que, si asumiéramos una postura como la utilitaria, según nuestro autor, se trataría de: 1) Un *evaluador débil*, 2) Una deliberación concebida como una operación de *cálculo*, y 3) El uso de un lenguaje *neutro, observacional*, cercano a lo *cuantitativo* y, por contra, purgado de expresiones cualitativas y valorativas.¹⁸⁵ Mientras que, si nos situáramos del lado de Taylor, se trataría, en cambio, de 1) Un *evaluador fuerte*, 2) Una deliberación que se apoya en evaluaciones fuertes, las cuales, a su vez, requieren de: 3) Un cierto uso del lenguaje que, según el autor, permite que las opciones a elegir se describan apelándose a términos valorativos y cualitativos que Taylor considera “contrastativos”¹⁸⁶ (*contrastive*), v.g. “cobarde” como opuesto a lo “valiente”, “noble” como opuesto a “bajo”. Esta forma de hablar, de la que se ve necesitado el evaluador fuerte, según Taylor sería inasimilable al modo de expresarse que caracterizaría a un tipo de evaluador que sólo juzgue sus alternativas en términos desprovistos de carga valorativa, y entre las que, a diferencia de la manera en que son descritas por el evaluador fuerte, no tendría por qué darse una relación de oposición, de contraste, tal y como la que habría, a criterio del autor canadiense, entre las opciones a elegir que el evaluador fuerte caracteriza mediante apelativos que indican dicha oposición.

Recordemos que para Taylor dicha oposición *no* indica una relación de incompatibilidad meramente *accidental* o circunstancial, tal y como la que había, en los casos arriba comentados, entre “ir a nadar” e “ir a comer”, “ir a la playa” e “ir a la montaña”. A diferencia de lo que estos ejemplos ilustran, la incompatibilidad que expresa el evaluador fuerte mediante su lenguaje “contrastativo” es de carácter esencial, dado que para el agente las opciones en juego se oponen radicalmente entre sí porque

¹⁸⁵ De nuevo, le recuerdo al lector lo que, en un sentido bastante similar, se comentó en el primer capítulo con relación al modelo de agente, de deliberación práctica y de uso del lenguaje que se le puede atribuir a Gauthier.

¹⁸⁶ Una vez más, me temo que estaré utilizando una expresión no canónica en nuestra lengua, pero que creo captura la idea del autor.

éste las asocia a *modos de ser*, a *modos de vivir* a los que la persona considera inconciliables, y a los que no podría caracterizar si no utilizara un lenguaje valorativo que le permitiera describirlos de manera “contrastativa”. Tal es el caso, según los ejemplos comentados, del sujeto que elige cierta acción porque la considera “valiente” y, en cambio, descarta su alternativa precisamente porque la ve como un acto de “cobardía”. En otras palabras, estamos ante un tipo de agente al que Taylor se refiere como “evaluador fuerte”, y al que opone la figura de un “evaluador débil”. Al primero, el autor lo asocia con el uso lingüístico de “contrastos fuertes” porque, según nuestro filósofo, no habría manera de dar cuenta de una opción a la que se describe, por ejemplo, en términos de “valentía/coraje”, si no se tiene en mente su relación de oposición esencial a otra alternativa articulada mediante el uso de la expresión “cobardía”.

For strong evaluation deploys a language of evaluative distinction in which different desires are described as noble or base . . . courageous or cowardly . . . But this means they are characterized contrastively. Each concept of one of the above pairs can only be understood in relation to the other. No one can have an idea what courage is unless he knows what cowardice is, just as no one can have a notion of “red”, say, without some other colour terms with it contrasts. It is essential to both “red” and “courage” that we understand with what they are contrasts (1985c, p. 19).

Si bien es cierto que el evaluador fuerte, según Taylor, hace uso de un lenguaje mucho más rico en contrastes y en matices, un lenguaje que pinta al mundo en términos de cualidades y valores irreductibles entre sí, e irreductibles a cualquier unidad de medida, esto no significa, sin embargo, que siempre nos comportemos como tales evaluadores fuertes, tal y como arriba se mencionó. Y lo que es más importante, tampoco implica que cuando evaluemos débilmente entonces describamos nuestras alternativas de elección en términos no cualitativos, o mejor, en términos estrictamente *cuantitativos*. De modo que, para Taylor, tratándose *también* de las evaluaciones *débiles* resulta artificioso el intento –atribuido por nuestro autor al utilitarismo y a la “familia” de teorías a la que éste, según él, pertenece– de establecer como significativas únicamente aquellas descripciones que se ofrezcan en un lenguaje *cuantitativo* que permitiese la conmensurabilidad y, por tanto, el cálculo. Aunque es cierto que finalmente la elección entre opciones a las que se articula mediante el lenguaje propio de las evaluaciones débiles, es una decisión que se justifica apelándose a expresiones

tales como “he decidido hacer A en vez de B porque *me apetece más*”, o “me parece *más atractivo* A que B”, de todos modos hay en juego cualidades incomensurables; pues ¿cómo podríamos calcular *qué tanto* más deseable nos puede resultar un alimento dulce frente a otro salado, o cómo podríamos medir con la misma vara las bondades de irnos a la playa y las de irnos a la montaña? Para Taylor, en *todos* nuestros casos de decisión, trátase de cuestiones que tocan aquellos asuntos que asociamos a formas de vida valoradas “fuertemente”, o trátase de temas que *no* relacionamos con cuestiones de *valor* que tengan implicaciones para lo que *somos como personas* o para el *tipo de vida* en el que quisiéramos participar y que, por tanto, son objeto de evaluaciones “débiles”, las alternativas de decisión *no* podrían ser adecuadamente descritas desde un punto de vista *cuantitativo*. Es decir, ignorándose la gran variedad de cualidades que se ponderan y que son irreducibles a unidades de cálculo.

[. . .] weak evaluations are not quantitative either. That is, the alternatives cannot necessary be expressed in some common units of calculation and in this sense rendered commensurable. This has often been obscured by the recurring ambition of our rationalist civilization to turn practical reflection as much as possible into calculation, an ambition whose major expression has been the doctrine of utilitarianism (1985c, p. 17).

A los ojos de Taylor resulta, entonces, implausible y artificioso el proyecto reduccionista de intentar expulsar el lenguaje cualitativo del repertorio de las expresiones significativas,¹⁸⁷ reemplazándolo por un lenguaje meramente cuantitativo que permita hacer conmensurables las opciones de elección y, por tanto, convertir los juicios, las deliberaciones y decisiones humanas en una suerte de cálculo. Ahora bien, a pesar de que, tanto las evaluaciones débiles, como las fuertes, no puedan llevarse a cabo a la manera de un cálculo, habría que insistir en aquello que las distingue: la cuestión del valor. De ello da cuenta el tipo de expresiones que utilizamos cuando hacemos evaluaciones débiles: aunque podamos hacer una larga disquisición sobre por qué iremos a la playa este fin de semana y, en cambio ya no iremos a la montaña, o sobre por qué nos inclinamos más por un alimento dulce que por otro salado, en último término, para Taylor, no consideramos que estas opciones tengan un valor especial que

¹⁸⁷ Recuérdese el proyecto que, en este sentido, hizo suyo el positivismo lógico, mediante su búsqueda de un lenguaje puramente observacional, tal y como ya se comentó a propósito de los parentescos teóricos atribuibles a Gauthier.

justifique nuestra decisión como algo que añada o reste algún elemento de especial significado para el tipo de vida que queremos vivir o para el tipo de persona que quisiéramos ser. Pues, finalmente, si escogemos comernos algo dulce en vez de algo salado, o decidimos irnos al mar en vez de ir a la montaña, lo hacemos “porque sí”, porque así nos apetece. De manera que, aunque en las evaluaciones débiles tampoco se dan la conmensurabilidad ni el cálculo, es innegable que cuando nos comportamos como evaluadores débiles llegamos, finalmente, según nuestro autor, a la *experiencia inarticulada* de que algo nos apetece “porque sí”. De allí también el paréntesis que me permití introducir a propósito de Frankfurt y del posible mutismo al que quedaría reducido su agente modélico si dicho autor se empeña en que éste, a pesar de la naturaleza nada trivial que puedan tener sus preocupaciones, no pueda dar razón de éstas.

Este tema de la inarticulación en la que terminamos cuando evaluamos débilmente, tema que, como veremos más adelante, tiene importantes consecuencias, nos pone de presente otro elemento que habría que resaltar: tratándose de este tipo de evaluaciones no tendría mucho sentido que se nos pidiera *dar razones* más allá del “así me apetece”. De allí que el lenguaje que desplegamos cuando evaluamos débilmente nos alcanzaría para dar cuenta de características cualitativamente menos ricas, tales como “más atractivo”, “menos agradable”. En cambio, nuestra experiencia como evaluadores fuertes está ligada al intento de *dar razones* y a la necesidad de *articular* en un lenguaje más rico ciertas cualidades y contrastes fuertes, tales como los que señalamos cuando hablamos de “ser tacaños” o “ser generosos”, “cobardes” o “valientes”, “nobles” o “bajos”. Este sería el lenguaje que utilizamos cuando tratamos de *articular por qué* preferiríamos no actuar movidos, por ejemplo, por la avaricia o la cobardía, y en cambio preferiríamos hacerlo movidos por la generosidad o el coraje. Para Taylor, esta diferencia referida al lenguaje al que apelamos para articular nuestras experiencias como evaluadores fuertes, en franco contraste con el lenguaje que desplegamos como evaluadores débiles es, pues, indesligable del problema del *dar razones*, las cuales también son razones para la decisión. Sobre este último asunto, que nos lleva también al problema de las decisiones difíciles y los dilemas morales, volveremos más adelante. A continuación me detendré un poco más en el tema que, creo, es el que jalona toda la discusión de Taylor en contra del utilitarismo: su modelo

de agencia humana, el tipo de agente que podríamos ser si cumpliéramos con los requerimientos de un discurso como el criticado por nuestro autor.

2. Las formas de autointerpretación y la pregunta por el “yo” (*self*) que creemos ser

Si, como hemos visto, se hacen indesligables entre sí el modelo de agente que se quiere privilegiar, la forma de deliberación que se asume como definitoria de la capacidad reflexiva y de la razonabilidad atribuibles a dicho agente, y el lenguaje que se considera el adecuado para dar cuenta de tales deliberaciones, para Taylor esto nos lleva a plantearnos la siguiente cuestión: dada nuestra experiencia como agentes que reflexionan, juzgan, deliberan, deciden e incluso pueden exponer sus razones para actuar, una pintura de nosotros mismos en la que se nos muestre como siendo fundamentalmente evaluadores débiles o como siendo capaces de ofrecer razones para la acción únicamente sobre la base de evaluaciones débiles, ¿sería aquella con la que estemos dispuestos a identificarnos? Pienso que lo que Taylor tal vez quiera mostrarnos es que la figura del evaluador débil, aquel que sólo alcanza a caracterizar sus alternativas de decisión en términos como “más satisfactorio”, “menos divertido” o “me apetece más”; o aquel que sólo logra establecer las consecuencias que esas alternativas traerían para un supuesto bienestar que sea mensurable u objeto de cálculo, sería una pintura que dejaría por fuera aquello que acaso consideramos como lo más importante y definitivo de nuestra agencia humana, del tipo de agente o de “yo” (*self*) con el que quisiéramos identificarnos.

Taylor trata de mostrar que el modelo de evaluador que nos presenta un discurso como el utilitarista no sería plausible como tal modelo, en lo que se refiere al problema de cómo estaríamos dispuestos a pensarnos a nosotros mismos en tanto que agentes que evalúan sus deseos, o sus alternativas a elegir, y toman decisiones sobre la base de tales evaluaciones. Lo cual no tendría por qué obedecer, según nuestro autor, a que no nos sea posible hacer evaluaciones débiles, ni a que no podamos comportarnos como evaluadores débiles. De hecho, bien puede ocurrir que este sea el tipo de postura que tengamos que asumir la mayor parte de las veces, sobre todo en el caso –también posible, incluso nada infrecuente, aunque preocupante– de aquellas personas cuyos

únicos problemas parezcan ser dónde pasar el próximo fin de semana, o qué merendar esta tarde. Para Taylor, la implausibilidad de esta figura del evaluador débil reside en el tipo de agencia humana que propone: alguien que sólo puede hacer un registro de sus preferencias y calcular las consecuencias de sus alternativas de decisión para un bienestar que, a su vez, sólo puede ser descrito en términos de “mayor” o “menor satisfacción”.

2.1 Ser un evaluador fuerte o ser un simple “sopesador” (*weigher*) de alternativas

Según Taylor, la idea de un agente que sólo pueda juzgar y comportarse como un evaluador débil sería una posibilidad antropológica muy difícil de concebir. Por el contrario, para el autor, habría buenas razones para que nos tomáramos más en serio la figura del evaluador fuerte, en cuanto al tipo de modelo que dicha figura puede ofrecernos para pensarnos como agentes humanos, para pensar en qué podría consistir el hecho de que evaluemos nuestros deseos y nuestras opciones de elección, intentemos actuar por razones, deliberemos sobre estas razones y contemos con la capacidad para dar cuenta de ellas. Pues estas posibilidades, que tal vez no quisiéramos negarle a la agencia humana, cree Taylor que seguramente irían más allá de la habilidad para ofrecer un mero “registro” de nuestras preferencias o para calcular las consecuencias de nuestras alternativas de decisión en lo que se refiere, repito, a la obtención de una “mayor/menor satisfacción”. Por supuesto que no sería sensato pensarnos a nosotros mismos como careciendo de este último tipo de habilidades y posibilidades necesarias para la evaluación débil y el cálculo maximizador, entre otras razones, por la innegable importancia que revisten y la necesidad que de ellas tenemos. El problema es si, además de dichas habilidades y posibilidades, también quisiéramos y podríamos pensarnos como poseyendo las capacidades de los evaluadores fuertes. Lo cual, en mi opinión, nos llevaría de nuevo a un tema ya examinado por Frankfurt: la necesidad y los alcances de una evaluación de nuestros deseos, motivaciones y preferencias, así como la naturaleza misma de esta evaluación, asuntos que, como arriba se anunció, también podrían traer importantes consecuencias al momento de que se intente determinar las posibilidades de la razonabilidad que podríamos querer atribuirnos como agentes humanos.

Creo que para encontrar en Taylor algunas claves importantes para responder a esta última cuestión, habría que detenerse primero en aquello que para él es lo que está en juego; aquello que pienso que es lo que nos haría preocuparnos también por el asunto que perseguimos en la exposición sobre Frankfurt. A saber, la concepción que tenemos de *nosotros mismos* como agentes, como sujetos de juicio, reflexión y deliberación. Taylor cree que las dos mencionadas formas de evaluación, débil y fuerte, deberían sobre todo ser vistas como expresivas de dos modelos del agente humano, del yo (*self*): “The point of introducing the distinction between strong and weak evaluation is to contrast the different kinds of self that it involves” (1985c, p. 23). A un sujeto que únicamente hiciera evaluaciones débiles sólo le podríamos concebir decidiendo entre cuestiones tales como las que menciono arriba: pasar el fin de semana en la montaña o pasarlo en la playa, almorzar ahora o ir a nadar después, merendar algo dulce o comer algo salado. Taylor se refiere a este tipo de agente como un “simple sopesador de alternativas” (*simple alternatives weighter*) y frente a él opone su modelo, al que considera más plausible y más expresivo de esas capacidades reflexivas que atribuimos a nuestra especial forma de agencia humana: el evaluador fuerte. El primero se ocuparía, insisto, solamente en calcular el “peso relativo” que tendrían sus distintos deseos y las consecuencias que traería la satisfacción de éstos en lo que respecta a un supuesto bienestar que, a su vez, sería objeto de cálculo y, lo que es más importante, no podría ser articulado en términos que vayan más allá de expresiones tales como “así me apetece”, “me gusta más”, “me resulta más atractivo”. Este lenguaje y esta forma de deliberación se relacionan de manera importante con otro elemento que, para nuestro autor, tendría que llamar nuestra atención: una de las características más interesantes de este tipo de agente es que no se le podría atribuir cierta capacidad para “profundizar”, para reflexionar sobre sí mismo, capacidad que, por contra, sí cabría atribuírsele al evaluador fuerte.

[. . .] the strong evaluator has articulacy and depth which the simple weigher lacks. He has, one might say, articulacy about depth. But where there is articulacy, there is the possibility of a plurality of visions which there was not before. The simple weigher may hesitate, as before the *éclair* and *mille feuilles* . . . but we would not say that he envisages his situation of choice now one way now another (1985c, p. 26).

Al evaluador fuerte, según Taylor, por el contrario de lo que sucede con el evaluador débil, sí podríamos pensarlo como categorizando sus alternativas relacionándolas con un elemento de reflexión más complejo, el cual requiere de un lenguaje más rico para ser aprehendido, a saber, aquello a lo que arriba ya hemos apuntado: las *formas de vida* dentro de las cuales se inscriben sus posibles deseos. Si se vive la experiencia de un evaluador fuerte, dichos deseos no serían aprobados ni rechazados sin mediación, sino que este tipo de juicio implicaría una reflexión sobre el *tipo de persona* al que el evaluador fuerte quisiera parecerse y el modo de vivir en el que quisiera participar como tal persona. De ahí que sus elecciones, según Taylor, indiquen una capacidad para la reflexión sobre sí mismo, para profundizar en aspectos que no serían objeto de preocupación en el caso del simple sopesador. Para nuestro autor, tales aspectos a los que atiende el evaluador fuerte serían vistos por éste como elementos *centrales* de su propia *identidad*. No se trataría de aspectos que el sujeto considere como inesenciales o *periféricos* en lo que respecta a dicha identidad, aspectos tales como los implicados en las disyuntivas arriba mencionadas: ir a comer o ir a nadar, viajar el próximo fin de semana a la playa o mejor hacerlo a la montaña.¹⁸⁸ Por contra, según Taylor, las elecciones del evaluador fuerte apuntan más directamente a la manera en la que el agente se interpreta a *sí mismo*.

En este punto, convendría acudir a un ejemplo mediante el cual nuestro autor intenta ilustrar esta idea de las diversas formas de autointerpretarnos, relacionadas, a su vez, con el tipo de lenguaje que estemos dispuestos a utilizar para articular dichas autointerpretaciones. Taylor se refiere al caso de una persona que sufre de sobrepeso a causa de su adicción a los dulces. Según el autor, esta persona bien podría, si asume la postura de un evaluador fuerte, contarnos su problema haciendo uso de expresiones valorativamente cargadas, como “falta de autocontrol”, “degradación”, a las que opondría “dominio de uno mismo” o “dignidad”. Pero también puede ocurrir que la persona sea inducida, con miras a solucionar su problema, a utilizar otro tipo de términos para hablar de éste, términos vaciados de tal carga valorativa y más cercanos al lenguaje de un evaluador débil, es decir, expresiones parecidas a “mayor estado de bienestar”,

¹⁸⁸ Sin embargo, en este punto habría que hacer una salvedad: como veremos en breve, Taylor cree que *también* nuestras evaluaciones *débiles* requieren, como condición de posibilidad, de nuestros compromisos valorativos más fundamentales, los cuales serían el objeto de preocupación del evaluador fuerte.

“más satisfacción”, “menos colesterol”, “mejor estado físico”. Es posible, por tanto, que alguien que se empeñe en negar, en contra de la tesis de Taylor, la necesidad de hacer uso de términos valorativos fuertes y, en general, de hacer evaluaciones fuertes, insista en que tal vez a esta persona le sería de mayor utilidad, para solucionar su problema, abordarlo de manera “distanciada” o “clínica” y evitar entenderlo en dichos términos valorativos y, por contra, tomar la postura propia de un sopesador que busca lograr un estado de bienestar.

Ante este reclamo, nuestro autor estaría de acuerdo en que podría ser útil el cambio de perspectiva que asuma quien deje de ver su problema como una adicción que le resta control sobre su propia persona y comience, entonces, a considerarlo más bien como un asunto que atañe a factores tales como sus niveles de colesterol, su necesidad de sentirse bien, la posibilidad de subir escaleras sin experimentar cansancio, o de usar el tipo de ropa que más le gusta, o de asegurar una vejez con menos complicaciones, o de ahorrarse el dinero que le costaría enfermarse a causa de su obesidad. Sin embargo, Taylor insiste en que el agente no experimenta su situación del mismo modo si interpreta su problema de una forma tal que entren en consideración el tipo de actitudes que quisiera tener, o el tipo de persona que quisiera ser, por ejemplo, un individuo con un mayor dominio sobre sí mismo, que si lo interpretara de la otra manera. Es decir, si su modo de entenderlo no involucrara este tipo de cuestiones valorativas fuertes, sino sólo aquellas otras que, para ser articuladas, no requerirían de evaluaciones fuertes. Incluso, afirma nuestro autor, esta diferencia en la forma de *experimentar* la propia situación, atada a la forma de *autointerpretarse*, se vería como insoslayable si se repara en cómo asumirían su problema respectivamente el evaluador fuerte y el débil, si saliera al mercado un nuevo medicamento que les permitiera comer todos los dulces que se les antojaran sin ganar peso y sin enfermar por ello. En este caso, alguien que se autointerprete apelando a evaluaciones débiles bien podría darse todo el gusto que quisiera y echar mano del medicamento, sin experimentar su situación al modo en que lo haría alguien que apele a evaluaciones fuertes. Pero si la persona insiste en ver la cuestión en términos valorativos fuertes, entonces la existencia del medicamento no le impediría seguir empeñada en el asunto de su autocontrol, de evitar lo que ella podría considerar como indigno o degradante en la pérdida de éste.

Estamos, pues, ante dos formas posibles de autointerpretación, ante las cuales, nuestro autor se pregunta, ¿cuál de las dos es la más acertada? La respuesta de Taylor sería: eso depende de cuál de las dos sea para el agente aquella que éste experimente como la más reveladora de sí mismo, la menos ilusionante. Lo cual no indica aún que la actitud del evaluador fuerte tenga que ser reemplazada necesariamente por la del evaluador débil, ni tratándose de este ejemplo concreto ni, en general, para el agente que experimente la necesidad de establecer cuáles son esos referentes que supone en sus juicios sobre el tipo de persona a la que quisiera parecerse y el tipo de vida que quisiera vivir.¹⁸⁹ De nuevo, y retomando lo arriba comentado, pienso que en este punto se puede apreciar en Taylor la posibilidad de una respuesta al tipo de reclamos que autores como Watson, Bratman y Scanlon le hacen a Frankfurt, reclamos a los que se hizo referencia en el capítulo anterior. Pues creo que si se insiste en la plausibilidad de esta figura del evaluador fuerte, en sus posibilidades para la reflexión, el juicio, la deliberación y el dar razones, posibilidades que no se entenderían sin su vinculación con aquello a lo que el autor canadiense se refiere como “forma de vida”, entonces creo que podría entenderse/verse mejor el contexto en el que el agente evalúa sus deseos. Por tanto, pienso que podría apreciarse de manera más clara de dónde surgen, cómo surgen, cómo se articulan las razones por las cuales el sujeto lleva a cabo tal evaluación, y por qué lo hace de esta o de aquella otra manera.

Lo anterior, a mi entender, nos muestra que cobraría sentido algo que no podría lograrse si se insistiera en pensar al agente humano únicamente como capaz de evaluar en un sentido “débil” o si, como hace Frankfurt, se insiste en un agente modélico que no puede dar razones, ni se involucra en deliberaciones, ni juzga a partir de cuestiones valorativas ni normativas; y ello a pesar de que su importante noción del *caring* no apunta a dicho tipo de evaluación débil y denota, por el contrario, una capacidad autoevaluativa bastante cercana a la del evaluador fuerte de Taylor. Si, en contraste con este mutismo del agente de Frankfurt, pero sobre todo, si en un contraste aún más evidente tratándose del modelo del evaluador débil que Taylor atribuye al utilitarismo,

¹⁸⁹ Esto creo que nos mostraría, a los ojos de Taylor, por qué no bastaba con la diferencia entre *deseos de primer* y *de segundo orden*, sino que el mismo Frankfurt tuvo que acudir a una distinción adicional: la que habría entre los deseos y las *voliciones de segundo orden*, según lo que vimos en el capítulo anterior. Taylor diría que en el ejemplo de las dos personas que quieren combatir su adicción a los dulces y su obesidad, ambas tienen el deseo de segundo orden de no desear comer dulces, pero la diferencia estriba en las *evaluaciones* que ellas llevan a cabo, la una de manera débil, la otra de manera fuerte.

se asume la preeminencia que tiene la figura del evaluador fuerte en el discurso del canadiense, creo que los deseos evaluados y los posibles metadeseos desde los cuales se los valore y se les otorgue un papel en la *vida* de una persona, tendrían un sentido que, en su articulación como tal sentido, también podría ser compartido por *otros* agentes. Podrían, entonces, articularse, ofrecerse y entenderse las *razones* por las cuales el agente se identifica o se separa de ciertos deseos. En este punto estamos, pues, en mi opinión, ciertamente lejos del anticognitivismo y de la aporía a los que llegamos con Frankfurt, así como de la inarticulación o del mero lenguaje cuantitativo al que llegaríamos con el modelo del simple sopesador y, acaso aún más lejos, de las “razones del autointerés” a las que apelaría el agente egoísta de Gauthier.

De este modo, si se asume el modelo del evaluador fuerte, pienso que podrían aprovecharse mejor los aportes que, en mi opinión, hacen que la “persona” de Frankfurt represente un modelo mucho más plausible que el agente restringido de Gauthier y del utilitarismo, tal y como este último es visto por Taylor. Así, podría verse mejor el papel que jugarían los metadeseos y motivaciones de orden superior, las preocupaciones (el *caring*) a las que Frankfurt se refiere, ya que sería posible atarlas a esa suerte de *interpretación* que, según Taylor, el agente puede hacer de *sí mismo*, y de su *identidad*. Con lo cual, también se alcanzaría a explicar mejor, en respuesta a críticas como las de Watson, por qué evaluar los deseos, y por qué se produce en determinado momento la decisión de poner fin al proceso de tal evaluación mediante el compromiso definitivo que, según Frankfurt, llega a asumir el agente con un determinado deseo. De manera que dicha decisión no tendría por qué ser pensada como un acto de voluntad para el cual no haya razones, o como una elección que se llevaría a cabo *in abstracto*, por parte de un agente al que tampoco habría más remedio que pensar *in abstracto*, o sin vínculos, o sin contexto ni, por tanto, condenado al mutismo, a la inarticulación y al solipsismo. Antes bien, creo que si se asume la figura del evaluador fuerte, y se toma en serio su preocupación por una *forma de vida* entonces, desde Taylor, dichos actos de decisión/identificación podrían ser entendidos como obedeciendo a un sentido, y a unos vínculos que unirían al agente con esa forma de vida en la que quisiera participar; con su propio *proceso vital* evaluado a la luz de los compromisos anclados en dicha forma de vida, compromisos que, si se es un evaluador fuerte, y como veremos más adelante,

el agente está en capacidad de examinar, no solamente *in foro interno*, sino también de cara a otros agentes.

2.2 La persona y el *wanton* de Frankfurt; el evaluador fuerte y el evaluador débil de Taylor

Pienso que la figura del evaluador fuerte también permite hacer más claro el porqué de los reclamos hechos por Frankfurt, en el sentido de que la autointerpretación que lleva a cabo el agente no puede lograrse, según este autor, al modo en que la persona podría elaborar un relato de los eventos *externos* a ella. Creo que, desde Taylor, y teniendo en cuenta lo que llevamos hasta aquí, podría apreciarse mejor por qué este último tipo de relato no podría reemplazar la evaluación efectuada por el agente que está hablando de *sí mismo* e intenta articular el *sentido* de sus acciones y evaluaciones. De allí también que, en mi opinión, esta figura del evaluador fuerte podría ser de utilidad al objetivo de mostrar la razón de ser de las críticas lanzadas por Frankfurt en contra del modelo de la agencia humana a la que este autor descalifica como “fragmentada” y que, recordemos, le atribuye a cierta corriente de la moderna teoría de la acción. Me refiero al modelo según el cual, al decir de Frankfurt, a los agentes les “ocurre” que tienen deseos que pueden describir como eventos “externos” a ellos mismos, pero no como deseos que ellos “integran” a su historia como personas, y que hacen “suyos” asumiendo por ellos una responsabilidad, mediante su evaluación y el compromiso con tales deseos. Sin embargo, en este punto he de advertir que más adelante veremos por qué este “integrar en la propia vida” o “hacer-de-uno” algunos de los propios deseos, no tiene que significar, según Taylor, hacer una elección inexplicable, como parece querer decirnos Frankfurt, sobre todo si tenemos en mente sus nociones de necesidad volitiva y de lo impensable, así como su lectura de los casos y ejemplos mediante los que Frankfurt intenta ilustrar tales nociones.

Creo que este asunto nos lleva al centro de la argumentación de Taylor y nos hace retomar, a su vez, un motivo fundamental del discurso de Frankfurt en su crítica al modelo “fragmentado” del agente: el hecho de que la evaluación de los deseos y el preocuparnos por ciertos asuntos, sean experiencias insoslayables para nosotros como agentes humanos. Este mismo motivo pienso que en Taylor se traduce a la pregunta

sobre cuál de los dos modelos de *agente* evaluador, el débil o el fuerte, es el que nos parece más plausible para expresar lo que pensamos de nosotros mismos como agentes, o de nuestra experiencia como tales. Como arriba se comenta, el autor canadiense intenta mostrarnos que, aunque muchas veces nos comportemos como evaluadores débiles, no deberíamos dejar de pensarnos como siendo, en lo fundamental, evaluadores fuertes, a menos que quisiéramos pagar el precio de renunciar a incluir dentro de nuestra autoimagen aquellas características que solemos atribuir a un ser humano reflexivo y capaz de preocuparse por el tipo de persona que querría ser, o el tipo de vida que quisiera llevar. Lo cual sugiere, a mi entender, una analogía con la tesis de Frankfurt de que, si bien podemos comportarnos como *wantons*, tal vez no querríamos dejar de pensarnos como siendo “personas” capaces de preocuparse por ciertos asuntos a los que otorgan importancia incluso por encima de algunos de sus deseos, o seres capaces de preocuparse por la calidad de sus voliciones de orden superior.

En este punto, alguien podría objetarle a ambos autores que si hay un personaje típico de nuestra sociedad es precisamente el individuo que vive como un evaluador débil o un simple sopesador, personaje que también resultaría bastante cercano al *wanton* al que se refiere Frankfurt. Por tanto, se le podría alegar tanto a este último, como a Taylor, que el mencionado tipo de individuo “despreocupado” o “mero sopesador” no sólo no nos resulta descabellado o extraño como modelo de agencia humana, sino que, además, bien podría ser visto como un modelo bastante popular en el contexto de una sociedad como la nuestra. Sin embargo, creo que ambos autores podrían defender sus tesis alegando lo poco plausible que podría ser el que nos pensemos como ejerciendo ese tipo de agencia, la del evaluador débil, o la del *wanton*, de forma permanente y, sobre todo, como la condición de posibilidad de todas nuestras evaluaciones; especialmente, cuando las circunstancias de nuestra vida nos obligan a salir de la comodidad del simple sopesador. Esto último se intentará mostrar a continuación, en lo que se relaciona con la importancia que tiene para Taylor el hecho de que se nos presenten dilemas morales –término que, recordemos, no es del agrado de Frankfurt– y también en lo que respecta a su idea de que la deliberación que desarrollamos en casos de esa naturaleza supone de manera importante, para el autor canadiense, un horizonte valorativo que está implícito en los juicios de todo agente, incluso en los del evaluador débil.

3. Los dilemas morales y la imposibilidad del “yo desvinculado”¹⁹⁰

Pienso que para intentar convencernos de que haríamos bien en pensarnos fundamentalmente como evaluadores fuertes, Taylor ya ha dado los primeros pasos, haciéndonos plausible su pintura de este tipo de evaluador. Sin embargo, creo que aún le queda por encarar el tipo de críticas que insistan en excluir lo cognitivo de aquello que Taylor llama “valores” y “compromisos valorativos”. Si hiciésemos caso de dichas críticas, en mi opinión no tendríamos más remedio que aceptar, o bien una posición anticognitivista como aquella en la que, según se vio en el anterior capítulo, parece atrincherarse Frankfurt, o en una orilla distinta a la de este autor, pero también en contra de lo que creo que convencería a Taylor, considerar al cálculo maximizador como la única forma posible de “dar razones”. Oponiéndose a ambas opciones, pienso que Taylor intenta mostrar que, si se examina más de cerca la naturaleza misma del proceso que tendría lugar cuando un agente se ve compelido a aclararse sus compromisos valorativos más fundamentales, entonces puede apreciarse que el esfuerzo de la *articulación* de tales compromisos, así como el carácter *dialógico* de este proceso y la importante mediación que en ello juega el lenguaje, harían implausible la tesis de que los compromisos valorativos son el fruto de una elección irracional o inarticulable, de una suerte de *fiat* sobre el que no puedan ofrecerse razones (1985c, pp. 34 y ss.).

Para hacer plausible su postura, Taylor argumenta en contra de la tesis, atribuida por él a J. P. Sartre, de que la elección del propio proyecto de vida, que en Taylor asociamos al asunto de la propia identidad y de los compromisos valorativos de una persona es, finalmente, una elección “radical”, en el sentido de que para llevarla a cabo el agente no contaría con elementos de deliberación que le precedan. Al contrario de lo que nuestro autor pone en boca de Sartre, Taylor cree que dichos elementos serían necesarios como la condición para formular cualquier juicio o adelantar cualquier forma de deliberación, y se trataría de elementos que la persona no podría crear de la nada, al conformar ellos el contexto de referentes de valor o de sentido que le permitirían al agente apelar a razones.¹⁹¹ Creo que la discusión del canadiense se dirige en contra de

¹⁹⁰ Me permito traducir así las expresiones de Taylor, “disembodied ego” (1985c, pp. 35 y ss.) y “disengaged self” (1985d, introducción, pp. 1-12).

¹⁹¹ Taylor no sólo le atribuye esta tesis a Sartre, sino también a “una influyente escuela de filósofos anglosajones”, que creo que serían aquellos a los que podríamos considerar, o bien cercanos al

una postura anticognitivistica según la cual la persona que se encuentre en la situación de un dilema moral y se vea, por tanto, obligada a tomar una decisión difícil en algún momento álgido de su vida lo haría, en última término, en una suerte de vacío referencial que tornaría su decisión en aquello que Sartre llama una elección “radical”. Por esto, en lo que sigue intentaré mostrar de qué modo el alegato de Taylor en contra de esta posición sartreana igualmente tiene consecuencias que admiten aplicarlo al discurso de Frankfurt. Ello, justamente en razón del anticognitivismo que, según se vio, defendía este último autor, y que se apreciaba del modo más claro, y acaso en sus consecuencias más preocupantes, en la lectura que Frankfurt hacía de ejemplos como el del propio Sartre: el del muchacho que tiene que decidir entre permanecer al pie de su madre enferma o irse a defender su patria.¹⁹² Recordemos que Frankfurt interpretaba este y otros casos —como el de Lutero, y el de la madre que no se siente capaz de abandonar a su hijo— de un modo tal que, según él, los agentes involucrados en estas situaciones en las que éstos se vean abocados a tomar decisiones difíciles e incluso dolorosas, terminan por tomar su determinación final en razón de eso que les resulta “impensable” o que sencillamente no se sienten capaz de hacer, o de aquello que termine por “pesar más” al momento de decidir, y sin que en ello medien deliberaciones en las que se apele a referentes normativos, ni a razones. Así, la decisión de un agente que se enfrente a este tipo de dilemas, si nos situamos en el punto de vista de Frankfurt, pienso que sería vista como aquello que Sartre llama una elección “radical” y de allí que, a mi entender, la crítica de Taylor en contra de esta tesis que le atribuye al pensador francés, igualmente podría sernos de utilidad al objeto de obtener conclusiones sobre la apreciación que tendríamos que hacer sobre la postura de Frankfurt. Sobre todo en lo que se refiere a aquellos aspectos que, según lo ya comentado, resultan ser motivo de preocupación por parte de autores que, como Watson, Scanlon y Brantnam disienten de su anticognitivismo.

emotivismo de autores como Mackie y Stevenson, o bien a ciertas variantes del utilitarismo. Cf. Taylor (1985c, p. 29).

¹⁹² *El existencialismo es un humanismo*, pp. 28-33.

3.1 El problema de la elección “radical”

Para ilustrar lo que sucede en aquellas situaciones en las cuales nos vemos compelidos a aclararnos a nosotros mismos nuestros compromisos valorativos más fundamentales, esto es, aquellos que definen de manera eminente nuestra identidad, Taylor acude al conocido ejemplo tomado de Sartre, al cual ya se hizo referencia en el capítulo anterior. En contra de la conclusión que, según el primero, Sartre pretende obtener del ejemplo, el filósofo canadiense intentará mostrar que en una situación de esta naturaleza, aunque es cierto que el agente puede verse enfrentado a alternativas que bien admiten ser vistas como incomensurables –lo cual, como veremos, para Taylor justifica el uso del apelativo “radical”–, sin embargo la decisión final no será “radical” en el sentido de que su fundamento sea inarticulable y permanezca en el misterio, o más allá del alcance de las razones. El autor canadiense piensa que no se trata de una decisión en el “vacío” de una radical libertad que consista en no tener compromisos previos con ciertos referentes de valor o de sentido, en una suerte de creación de valores que consista en “elegir” o “crear” obligaciones posteriores a la situación del dilema. Lo cual implica también para nuestro autor, y a despecho de lo que le atribuye a Sartre, que la decisión del agente puede ser *articulada* y, por tanto, así mismo entraría bajo el dominio del dar razones.¹⁹³

Lo anterior se explica porque, para Taylor, y al contrario de lo que sugiere Sartre, el muchacho del ejemplo se encuentra ante un auténtico dilema *moral* precisamente dado que se enfrenta a *demandas (claims)* que considera justificadas, alternativas de elección a las que ve como auténticas *obligaciones*, en tanto que reclamos que reconoce como fundamentales para él. Justamente por tratarse de reclamos, de demandas que son objeto de juicio, de examen, y no de una “creación/invención”, no de una “elección”, y menos aun de una elección que se realice en el vacío, dichos reclamos no podrían ser el producto de la sola capacidad de elegir, o de una radical libertad que, según Taylor, Sartre le atribuye a todo agente humano. Para el canadiense, al deliberar, el muchacho del ejemplo tendría que examinar las razones

¹⁹³ En *El ser y la nada* (p. 590), Sartre se refiere a la elección del proyecto personal de vida, como una elección “absurda: “Esa elección no es absurda en el sentido en que, en un universo racional, surgiera un fenómeno que no estuviera en conexión con los demás; sino que es absurda en el sentido de que es aquello por lo cual todos los fundamentos y razones vienen al ser; aquello por lo cual la misma noción de absurdo recibe su sentido. Es absurda en cuanto allende a todas las razones”.

que pueden argüirse desde posiciones que encarnen a cada una de las demandas morales en disputa y estas razones mostrarían el valor y la forma de vida que, en cada caso, estarían en juego. Pero estas razones deben ser articuladas; o mejor, quien se encuentra ante un dilema de esta naturaleza siente que cada una de las opciones lo “llama”, o “tira de él” y es por eso que necesita articularlas: con el fin de aclararlas y ponderarlas. De este modo, para Taylor, la decisión se dará sobre la base de una deliberación que asume la forma de una confrontación entre las razones que se ofrezcan desde eso que el canadiense ve como los reclamos morales que el agente siente que justificadamente se le hacen –cuidar de su madre enferma, defender su patria–. En esta confrontación, que es una confrontación entre autointerpretaciones, tal como la ve Taylor, el muchacho del ejemplo de Sartre sólo podrá actuar de buena fe si, antes de optar por una de sus alternativas de elección, intenta examinarlas a la luz de aquello que resulte más importante para él. Empero esto último, a su vez, y por definición, a los ojos de Taylor no depende de una “creación” o de una elección previa por parte del chico. Éste intentaría, por el contrario, arribar a una toma de decisión, pero previo paso por el mencionado examen, y con miras a establecer cuál de las versiones en las que se articulan sus opciones a elegir, conforma aquella autointerpretación que se le revele como la más veraz, la menos ilusionante sobre sí mismo y sobre lo que realmente importa para él. Esto es, aquello a lo que el muchacho se descubra (a sí mismo) como sus fundamentalmente más “atado”, en el sentido de que se le revelen sus compromisos valorativos más importantes.

Lo que tal vez justifique el uso del adjetivo “radical” es, para Taylor, el hecho de que el agente debe llevar a cabo su proceso de deliberación a partir de la articulación de las autointerpretaciones en disputa pero, para ello, desde cada una de éstas se le ofrecería al sujeto un discurso en el que se muestran como *positivas*, obligatorias y significativas ciertas características de la forma de vida o de la trayectoria vital que dicho discurso aconseje, al tiempo que, desde éste se hace una valoración *negativa* de la forma de vida que se asocia al discurso contrario (o a la autointerpretación contraria). De este modo, la elección entre tales autointerpretaciones sí se parecería a una elección entre inconmensurables y, en ese sentido, podría ser considerada como “radical”. Recordemos aquí lo que ya se ha comentado sobre el lenguaje “contrastativo” en el que, según Taylor, dichas autointerpretaciones tendrían que ser articuladas por parte del

evaluador fuerte. Lo cual nos hace ver que no cabría hablar de conmensurabilidad entre los relatos así enfrentados y, por tanto, que la elección entre ellos, o entre lo que cada uno de ellos aconseje, sería una elección que admitiría ser calificada de “radical”. Empero, según Taylor, esta escisión e incomensurabilidad de las autointerpretaciones del agente no puede soslayar que el hecho mismo de que se nos presenten dilemas morales auténticos, como el del ejemplo, sería inexplicable si, como afirma Sartre, nuestros compromisos valorativos obedecieran a una elección radical, a una elección que se diera antes de que la persona haya establecido lazos con tales compromisos, o antes de que se halle vinculada con un trasfondo de referentes de valor y de sentido. Por el contrario, según el canadiense, justamente una situación en la que se nos plantea un dilema moral requiere como su condición dicho trasfondo u horizonte y es por eso que, al vernos involucrados en tal situación, nos sentimos obligados a intentar aclararnos los compromisos que suponen dicho horizonte; lo cual no sería, entonces, equiparable a una “elección” de tales compromisos.

No obstante, habría que advertir que, acerca de estos últimos, no siempre ni necesariamente tenemos por qué tener una tal claridad, ni mucho menos que ésta anteceda a la deliberación que se llevaría a cabo con motivo del dilema, entre otras razones, por la naturaleza del mencionado trasfondo u horizonte referencial que se torna en problemático ante una situación de este tipo. Por tanto, la experiencia misma del dilema, el sentirse llamado por las demandas contradictorias que en él se le plantean al agente, demandas que, como tales, son vividas por éste como exigencias a las que quisiera responder, dado que las ve como justificadas, es una experiencia que le obliga a sacar a la luz, mediante el lenguaje, eso que es objeto de tales demandas. Es decir, esta experiencia hace que la persona sienta que se le exige llevar a cabo un proceso de articulación, de explicitación y deliberación sobre aquello que sea lo más valioso para ella, pero que también resulta ser aquello sobre lo que posiblemente el sujeto no tenía plena conciencia de que revestía tal importancia para su vida.

A los ojos de Taylor esto tendría importantes implicaciones. En *primer* lugar, la deliberación llevada a cabo por quien se enfrenta a un dilema moral consiste en un esfuerzo por *articular* algo, por lograr que el lenguaje saque a la luz “algo” que la persona no ha elegido ni creado, sino que se le “revela” y que, por tanto, tiene para ella un cierto carácter de objetividad que le hace irreducible a un mero fruto de su capacidad

para elegir o crear. De allí que, repito, el agente experimente la necesidad de articular sus opciones de elección al modo de relatos, de autointerpretaciones que pugnan entre sí, dado que solamente de este modo la persona puede intentar hallar una solución que, a su vez, dependerá de cuál de esos relatos termine por mostrársele como aquel que le lleve a una conciencia más clara, menos engañosa, de sus evaluaciones fundamentales. Lo cual implica, en *segundo* término, que estaríamos hablando de un esfuerzo por explicitar, por dar razones; por hacer que aquello que en principio parecía encontrarse en el límite de lo no articulado, pueda salir a la luz y sea desplegado gracias al lenguaje. Se trata entonces de un esfuerzo bastante alejado del “fiat”, o del “absurdo”, y que mal podría ser equiparado a lo meramente “irracional”. Finalmente, y en *tercer* lugar, el sujeto que lleva a cabo este esfuerzo de articulación, si bien en algunas ocasiones puede deliberar en soledad, sin embargo va a hacer uso de un lenguaje que lo vincula con *otros*; más exactamente, con aquellos otros de los cuales (y con los cuales) aprendió el uso de ese lenguaje, y de entre ellos, esos “otros” que acaso hayan resultado más definitivos para su autocomprensión como sujeto. En este sentido, la articulación de esta autocomprensión no es un monólogo *sensu stricto*, sino un diálogo, unas veces con otros realmente presentes, y otras veces introyectando en uno mismo a esos otros. El lenguaje en el cual articulamos nuestros compromisos valorativos nos vincula, por tanto, con nuestros “otros” y con lo que también tiene significado y es fundamental para ellos.¹⁹⁴ En este ámbito, como en todos, no funcionan, pues, los lenguajes privados, lo cual además refuerza la tesis anterior acerca de que no podríamos situar el proceso de aclaración de las evaluaciones fundamentales en el dominio de lo irracional.

Lo anterior lleva a que, para el filósofo canadiense, la experiencia de sentirnos necesitados del logro de la mencionada “revelación”, que es más bien una aclaración o, mejor aun, una autoaclaración, tendría que ser descrita como la experiencia de la necesidad de ser fieles a “algo”, de no traicionar ese algo que, si *no* nos aclaramos,

¹⁹⁴ En *La ética de la autenticidad* (pp. 68-69), Taylor insiste en este aspecto lingüístico y dialógico: “Nadie adquiere por sí mismo los lenguajes necesarios para la autodefinición. Se nos introduce en ellos por medio de los intercambios con los otros que tienen importancia para nosotros, aquellos a los que G. H. Mead llamaba los *otros significativos*. La génesis de la mente humana es en este sentido no *monológica*, y no constituye algo que cada cual logre por sí mismo. . . Se espera que desarrollemos en una medida considerable nuestras propias opiniones, puntos de vista y actitudes hacia las cosas mediante la reflexión solitaria. Pero no es así como funcionan las cosas en el caso de las cuestiones importantes, como la definición de nuestra identidad. Esta queda definida siempre en diálogo, y a veces en lucha, con las identidades que nuestros otros significativos quieren reconocer en nosotros” (las cursivas son de Taylor).

podría ser así traicionado. Y es la preocupación ante este peligro lo que marca a un dilema moral con el componente de una duda dura e incómoda, e incluso, de una suerte de angustia que se experimenta dada la preocupación por la posibilidad de errar con respecto a “algo” que, si bien es de “nuestra” vida y en este sentido nos pertenece, también constituye “algo” a lo que quisiéramos apuntar con nuestras acciones y elecciones; por tanto, no se trataría de un objeto “externo” del que estemos desvinculados ni que podamos elegir a capricho. Esto, a su vez, implica que, de *un lado*, las autointerpretaciones que se esfuerza por articular una persona presentan una naturaleza completamente distinta a la descripción que pueda hacer el agente de un objeto externo a ella/él mismo.

El llevar a cabo el primer tipo de esfuerzo cambia su propio “objeto” de interpretación, de manera que la experiencia de intentar articular una autointerpretación no dejará intacta a la persona que la lleva a cabo, ni el tipo de vivencias que pueda tener de sí misma y del mundo. Para Taylor, cada nuevo intento de autointerpretación transforma de tal modo a la persona, que abre ante ella ciertas posibilidades de experimentar su propia situación y de ver el mundo, al tiempo que le cierra o le oscurece otras formas posibles, que se tornan en poco significativas e incluso que se desacreditan ante sus ojos. Lo cual implica también un cambio en el tipo de lenguaje que le resulte significativo y “aclarador” al agente que se vale de tal lenguaje para intentar dar cuenta de sí mismo. No sería, pues, lo mismo equivocarse cuando se describe incorrectamente un objeto externo al agente, que equivocarse cuando se trata de autointerpretarse y explicitar aquello que constituye para uno sus compromisos valorativos más importantes.¹⁹⁵ De allí que, *en segundo lugar*, Taylor insista en que esta

¹⁹⁵ “This is a conflict of self-interpretations. Which one we adopt will partly shape the meanings things have for us. But the question can arise which is more valid, more faithful to reality. To be in error here is not just to make a misdescription, as when I describe a motor-vehicle as a car when it is really a truck. We think of misidentification here as in some sense distorting the reality concerned” (1985c, p. 22).

Más adelante afirma:

“These articulations are not simple descriptions, if we mean characterization of a fully independent object . . . articulations are attempts to formulate what is initially inchoate, or confused . . . but this kind of formulation or reformulation does not leave its object unchanged [. . .] our self-interpretations are partly constitutive of our experience. For an altered description of our motivation can be inseparable from a change in this motivation. But to assert this connection is not to put forward a causal hypothesis: it is not to say that we alter our descriptions and then as a result our experience of our predicament alters. Rather it is that certain modes of experience are not possible without certain self-descriptions. [. . .] certain descriptions of experience are unacceptable or incomprehensible to some people because of the nature of their experience [. . .] descriptions and experience are bound together in this constitutive relation that

forma de experiencia que conlleva el esfuerzo por sacar a la luz tales compromisos no sería posible sin las ataduras previas que tenemos con ese “algo” que se intenta aclarar mediante dicho esfuerzo, ataduras que nos conectan con unos referentes de sentido y de valor que, insisto, anteceden a un dilema moral como el del ejemplo de Sartre, y sin las cuales la situación descrita no tendría justamente el carácter de tal dilema. Dichas ataduras, que no son otras que aquellas que nos arraigan a nuestras evaluaciones más fundamentales incluso antes de que, repito, tengamos la suficiente claridad sobre tales evaluaciones, por definición se constituyen en lazos, en compromisos que nos impiden elegir radicalmente, o en el vacío, o en una suerte de libertad que consista en la desvinculación con cualquier referente que no obedezca a nuestra creación o voluntad.

Para Taylor, sin estos lazos sería imposible que un personaje como el muchacho del relato de Sartre pudiera sentirse auténticamente comprometido con las opciones a elegir que se le plantean y, por ende, tampoco podría experimentar su dilema como tal, ni sentir la carga de dolor que éste comporta. Si dichas opciones hubieran sido meramente elegidas por el chico del ejemplo, entonces él no experimentaría precisamente dicho dolor, no viviría su situación y su obligación de tener que escoger entre sus alternativas con la carga de angustia con que la vive.¹⁹⁶ A su vez, esta última sólo se experimenta ante un auténtico dilema moral justamente porque al agente no le resulta nada fácil lograr establecer, finalmente, cuál de los dos relatos acerca de lo que debe hacer es el más fiel, el más revelador, el menos “ilusionante” acerca de sus más fundamentales evaluaciones y de la acción que tendría que llevar a cabo si quiere ser fiel a dicho relato.

admits causal influence in both directions: it can sometimes allow us to alter experience by coming to fresh insight; but . . . it circumscribes insight throw the deeply embedded shape of experience for us. Because of this constitutive relation our descriptions of our motivations, and our attempts to formulate what we hold important are not simple descriptions in that their objects are not fully independent. And yet they are not simple arbitrary either, such that anything goes. There are more or less are adequate, more or less truthful [. . .] a new articulation does not leave its “object” evident of obscure to us in the same manner or degree as before . . . of shaping it makes it accessible and/or inaccessible in new ways. Now our articulations, just because they partly shape their objects, engage our responsibility in a way that simple descriptions do not . . . ” (pp. 36-38).

¹⁹⁶ Si se tratara de elegir o “crear” sus opciones, y de optar “radicalmente” sin que medien razones ancladas en la trayectoria vital y en el horizonte de sentido y de valor que antecede la situación del dilema, y que lo constituye como tal, entonces, según Taylor, el chico del ejemplo viviría su experiencia de verse obligado a escoger entre cuidar a su madre enferma o luchar por la liberación de su patria, de la misma manera que si se tratara, por ejemplo, de una disyuntiva entre velar por su madre o irse de vacaciones a la playa (1985c, p. 30).

Estas evaluaciones anclan en eso a lo que Taylor se refiere como el *horizonte* valorativo a partir del cual es posible que, como agentes, hagamos *cualquier* juicio sobre lo preferible; incluso, tratándose de nuestras evaluaciones débiles. El autor canadiense cree que en ausencia de este horizonte sería inconcebible que tomáramos decisión alguna, o hiciéramos juicio alguno. Pensarnos a nosotros mismos como careciendo de este horizonte sería, para Taylor, pensarnos como sujetos sin identidad, sin ningún tipo de referente a partir de cual elegir y juzgar, lo cual, amén de ser una situación sumamente difícil de pensar, estaría lejos de ser el paraíso de libertad y autosuficiencia que pudiera parecer a primera vista. Pues, según el autor, la posibilidad de elegir en el vacío, de determinar creativa y radicalmente eso que sería objeto de nuestras decisiones, dado que careceríamos de todo horizonte referencial, sería más bien la causa de la imposibilidad de ejercer una agencia propiamente humana o, si se quiere, la causa de la mayor desorientación que, lejos de hacernos “libres”, más bien equivaldría a una terrible pérdida de nosotros mismos. En este sentido, una vez más, Taylor ve en esta idea de “libertad” entendida como la desvinculación con todo horizonte de valor y de sentido, un engañoso espejismo producido por nuestra cultura, engaño que, no obstante, y como arriba se comentó, tiene relación con un ideal moral bastante entendible y muy caro para la modernidad: el de un sujeto que puede hacerse dueño de sí mismo por la vía de la falta de una identidad que le ate a las fuerzas de su mundo social, y como una forma de verse también liberado de las fuerzas del mundo físico, e incluso de sus propias carencias, vulnerabilidades y pasiones como sujeto.¹⁹⁷

¹⁹⁷ “Our identity is [therefore] defined by certain evaluations which are inseparable from ourselves as agents. Shorn of these we would cease to be ourselves . . . shorn of these we lose the very possibility of being an agent who evaluates; that our existence as persons, and hence our ability to adhere as persons to certain evaluations, would be impossible outside the horizon of these essential evaluations, that we would break down as persons . . . if I were forced by torture or brainwashing to abandon these convictions by which I define my identity, I would be shattered, I would be no longer be a subject capable of knowing where I stood and what the meanings of things were for me, I would suffer a terrifying breakdown of precisely those capacities which define a human agent . . . The notion of identity refers us to certain evaluations which are essential because they are the indispensable horizon of foundation out of which we reflect and evaluate as persons. To lose this horizon, or not to have found it, is indeed a terrifying experience of desegregation and loss. . . The agent of radical choice would at the moment of choice have ex hypothesi no horizon of evaluation. He would be utterly without identity. He would be a kind of extensionless point, a pure leap into the void. But such a thing is an impossibility or rather could only be the description of the most terrible mental alienation. The subject of radical choice is another avatar of the recurrent figure which our civilization aspires to realize, the disembodied ego, the subject who can objectify all beings, including his own, and choose in radical freedom. But this promised total self-possession would in fact be the most total self-loss” (1985c, pp. 34-35).

3.2 La libertad del agente *à la* Gauthier vs. las ataduras del evaluador fuerte. De la elección radical de Sartre a lo “impensable” de Frankfurt

En contra de esta imagen de un yo desvinculado del mundo y de otros yoes, para Taylor, los imprescindibles lazos que tiene el individuo con su horizonte valorativo, con las otras personas con las cuales comparte las claves de comprensión de tal horizonte, así como la naturaleza dialógica del proceso mediante el cual el sujeto asume el esfuerzo de dicha comprensión, indicarían que el agente en realidad no se encuentra “solo”, ni desvinculado de su mundo –ni físico, ni social–, no pudiendo, entonces, situarse “por fuera” de su proceso de deliberación-aclaración a la manera que podría hacerlo un observador neutral que se limitara a describir “hechos” en los que no esté involucrado. En este punto, creo que conviene recordar, de una parte, las protestas de Frankfurt en contra del modelo “fragmentado” de agente y, de otra, la tesis de Taylor acerca del ideal moral que explicaría, según él, esta imagen del yo-desvinculado: la idea de libertad asociada a este tipo de agente que, en virtud de su “desvinculación” del mundo, gozaría de un cierto “poder” para “controlar”, y de una cierta “asepsia”, que tal vez se echen de menos en el evaluador fuerte que está tratando de decidir frente a un dilema moral.

Pienso que el evaluador débil que está tomando una decisión de tipo técnico, en la que busca establecer el medio más eficiente para el logro de cierta finalidad a la que no pone en cuestión, en cierta forma puede “salirse”, liberarse de su situación y observarla desde fuera, como si tomara el lugar de *cualquier* otro agente, el cual llevaría a cabo una suerte de peritaje sobre el asunto. Es decir, estaríamos ante un sujeto que creo nos llevaría de vuelta al agente modélico de Gauthier, agente al que podríamos ver como una suerte de evaluador externo, neutral, incontaminado, que podría aconsejar a *cualquier* otro agente sobre la viabilidad técnica de las opciones a elegir, dado que las características relevantes de la situación son describibles independientemente del sujeto, esto es, de *quién* sea la persona que se encuentre frente a ellas. Lo cual tal vez nos recuerda los alcances que le atribuía Gauthier a su particular idea de “imparcialidad”. Por esta razón, pienso que un agente como el que aparece en *La moral por acuerdo*, al igual que aquel que Taylor critica en el discurso utilitarista, no sólo debe prescindir de un lenguaje valorativamente cargado, sino que, además, tendría la posibilidad de expresar en lenguajes formales las características de toda situación de elección. Así, a

esta última se la podría describir como un estado de cosas cuyas características “estructurales” pueden ser expresadas formalmente, en términos cuantitativos, y como una situación en la que puede hallarse “cualquier” agente pensado como “neutro”, como desvinculado de cualquier compromiso valorativo. Esto es así incluso si se supone previamente un marco de sus “preferencias”, pues éstas, si se asume una postura como la de Gauthier, serían sólo eso: un supuesto, unas condiciones iniciales que no son el objeto de deliberación. Ese mismo agente modélico presentado por este tipo de discurso puede, por tanto, verse liberado de ataduras, logrando así situarse por fuera de sus circunstancias, lo cual, después de todo, admite, como bien dice Taylor, ser visto como una atractiva posición de libertad, en la que se logran controlar los rasgos de la situación y en la que cabe hacer descripciones desapasionadas u objetivas de la misma.

Por contraste, creo que el agente pensado como un evaluador fuerte no gozaría de esta forma de libertad cuando lleva a cabo sus deliberaciones, sobre todo aquella en la que intenta establecer cuál de sus alternativas de decisión es la más fiel a lo que considera más valioso. Pues, en primer término, y como ya se ha dicho, se encuentra con el límite de que eso valioso no es objeto propiamente de “elección”, sino de aclaración o desvelamiento, lo cual indica que la posición del sujeto no es propiamente la de quien “controla”. Recordemos que el agente de Taylor no puede por sí solo determinar a capricho qué es lo significativo o lo fundamental, ya que se encuentra vinculado con su horizonte de valor y de sentido y, al intentar dar razón de tales vínculos, debe traerlos a la luz, articularlos y examinarlos mediante el uso de un lenguaje cuyos significados tampoco controla a voluntad, un lenguaje que debe compartir con otros agentes. Un sujeto al que se piensa de este modo no “elige”, entonces, qué es lo significativo, no escoge sin más sus compromisos con lo que considera valioso —como, recordemos, en cambio, sí podría elegir el evaluador débil que está intentando decidirse entre irse a nadar o almorzar primero—. Creo que tal “decisión” ni siquiera se produciría cuando un evaluador fuerte “decida” renegar de alguno de esos compromisos.¹⁹⁸ Pues, en cuanto le ocurriera algo así, se encontraría en

¹⁹⁸ “La autoelección tiene sentido sólo porque ciertas cuestiones son más significativas que otras. No podría pretender que me elijo a mí mismo y desplegar todo un vocabulario nietzscheano de autoconformación, sólo porque prefiero escoger un filete con patatas en vez de un guiso a la hora de comer. Y qué cuestiones son las significativas no es cosa que yo determine. Si fuera yo quien lo decidiera, ninguna cuestión sería significativa. Pero en ese caso, el ideal mismo de la autoelección como *ideal moral* sería imposible. De modo que el ideal de la autoelección supone que hay *otras* cuestiones significativas

una situación igual de complicada que la de Lutero, o la del muchacho del ejemplo de Sartre, ya que al renegar de un compromiso al que antes consideraba de la mayor importancia, tendría que mostrarse a sí mismo la preeminencia de los vínculos que lo atan a *otro* compromiso valorativo que se le “revele” como más fundamental –y acaso como opuesto– al compromiso del que ha decidido renegar. En este sentido, el agente que estuviera ante este tipo de decisión, aunque en ella parezca que hay un cambio radical que sugiera una suerte de “autocreación”, sin embargo también estaría atado, inmerso en su situación, y el tipo de “libertad” que se le pueda atribuir aparecería, no como una posición inicial de “control”, sino como un *logro* que sólo puede obtenerse al final, luego de haber desarrollado un honesto –y tal vez doloroso– esfuerzo de autoaclaración.

Incluso tratándose de este tipo de cambios radicales, pienso que Taylor nos diría que, al requerirse en ellos del esfuerzo del agente por lograr la lucidez, por articular o hacer explícito lo que le resulta más valioso, serían siempre esfuerzos y logros parciales, susceptibles de reevaluación y crítica, no sólo porque la persona no pueda situarse por fuera de su propia situación y, en este sentido, no logre “controlarla” a la manera de un “objeto” externo a ella, sino también, y aunque Taylor a veces no sea del todo enfático al decirlo, porque creo que el agente no puede ser pensado como teniendo un acceso sin mediaciones a esos compromisos valorativos en los que se supone reside su propia identidad y, de este modo, no sería transparente para sí mismo. En otros términos, pienso que la persona no tendría un acceso sin mediaciones a lo que ella misma es, dado que le sería imposible “salirse” de sí misma y observarse desapasionadamente y de manera prístina como si se mirara a través de un microscopio. Recordemos una vez más que, para hablar de sí mismo, el agente debe contar con la mediación del lenguaje, lo

más allá de la elección de uno mismo. La idea no podría persistir sola, porque requiere un horizonte de cuestiones de importancia, que ayuda a definir los aspectos en los que la autoconformación es significativa. Siguiendo a Nietzsche, soy ciertamente un gran filósofo si logro rehacer la tabla de valores. Pero esto significa redefinir los valores que atañen a cuestiones importantes, no confeccionar el nuevo menú de MacDonald's, o la moda en ropa de sport de la próxima temporada. El agente que busca significación a la vida, tratando de definirla, dándole sentido, ha de existir en un horizonte de cuestiones importantes [. . .] Sólo puedo definir mi identidad contra el trasfondo de aquellas cosas que tienen importancia. Pero poner entre paréntesis a la historia, la naturaleza, la sociedad, las exigencias de la solidaridad, todo salvo lo que encuentro en mí, significaría eliminar todos los candidatos que pugnan por lo que tiene importancia. Si existo en un mundo en el que la historia, o las exigencias de la naturaleza, o las necesidades de mi prójimo humano, o los deberes del ciudadano, o la llamada de Dios, o alguna otra cosa de este tenor tiene una importancia que es crucial, puedo yo definir una identidad para mí mismo que no sea trivial” (1994, pp. 74-75; las cursivas son de Taylor).

cual nos hace ver el peso que tendrían las voces de los otros agentes con los que comparta ese lenguaje. Así, las versiones que la persona tenga de sí misma también podrían ser siempre objeto de revisión; podrían estar más o menos claras, más o menos distorsionadas. Creo, entonces, que estamos ante la imagen de un ser que se encuentra “contaminado”, “atado”, lejos de la asepsia y la autosuficiencia del agente de las decisiones técnicas y, en cambio, se halla atado a sí mismo, a su situación, a los otros agentes con los que conversa sobre ella, al lenguaje en el que se realiza esa conversación y, por supuesto, a aquello que considera lo más valioso, aunque no lo tenga del todo claro. Y si la deliberación que llevase a cabo con el fin de lograr dicha claridad, como vimos, tampoco podría darse gracias al uso de un lenguaje valorativamente neutro, esto nos llevaría a excluir a la figura de un agente “abstracto” cuyo lugar pueda ocupar “cualquiera”, tal y como, en cambio, ocurriría en el caso del maximizador de Gauthier.

Si quien adelanta esta forma de deliberación, según Taylor, no puede decidir a capricho cuál de las alternativas de decisión encarna mejor sus compromisos valorativos, así como tampoco puede determinar cuáles han de ser dichos compromisos, sino que trata de aclararlos, creo que también cabe hacer alguna precisión sobre los aportes de Frankfurt, comentados en el capítulo anterior. Pienso que, desde Taylor, Frankfurt estaría en lo cierto al afirmar que las preocupaciones fundamentales de las personas (su *caring*) no son objeto de caprichosa elección e, incluso, algunas de ellas parece que se les imponen, y por eso ven en su contravención algo impensable para ellas. Sin embargo, tal vez Taylor pediría mayores explicaciones frente a la afirmación de Frankfurt de que lo que decida hacer el chico del ejemplo de Sartre dependerá de qué “pese más” en sus preocupaciones. Pues creo que, desde Taylor, esta idea de un mayor peso relativo de unas preocupaciones frente a otras, de suyo tendría que ser explicada por una cuestión de *sentido*, un sentido que nos remite a eso que arriba se denominó “forma de vida”, entendido como un referente complejo –al modo de una “red”– desde el que se explicaría lo que las cosas *significan* para las personas. Luego, la decisión del agente no tendría por qué ser vista como el mero hecho –bruto e inarticulable– de que “la alternativa A pesa más que la alternativa B”, tal y como, en cambio, sí sucedería ante la disyuntiva de merendar algo dulce o algo salado.

Si en una situación de dilema moral, como el caso presentado por Sartre, o el caso de Lutero presentado por Frankfurt, el agente evalúa su situación y sus opciones, y llega a concluir que para él “A pesa más que B”, creo que esto se deberá, en términos de Taylor, a que A resulta más *significativa*, más valiosa o más importante para el agente, dado que éste contrasta sus opciones de elección con un trasfondo de validez y sentido. La decisión final a la que llegue la persona, por tanto, es un punto de *llegada* antecedido por un *contexto*, de modo que no surge como un salto en el vacío, sino como el producto del esfuerzo hecho por el agente por articular sus opciones, por tratar de caracterizarlas lo más ricamente posible, esto es, por *dar razones*, lo cual nos haría ver la implausibilidad de la postura anticognitivista de Frankfurt.¹⁹⁹ Pienso que esto último explicaría precisamente la dificultad de esta particular forma y proceso de deliberación, y el “arte” que sería propio de ésta: a saber, el pintar fiel y detalladamente las opciones a elegir. Esto es, articular las *razones* para la decisión, razones ancladas en la autocomprensión del agente, en aquello que considera más fundamental y significativo para dar cuenta de sí mismo y de sus acciones, si bien eso fundamental-significativo no está del todo claro para él, y de allí que alcanzar una mayor claridad sea, por tanto, un logro.

Si ese horizonte valorativo que, para Taylor constituye la posibilidad de que juzguemos como agentes humanos incluso, repito, tratándose de nuestras evaluaciones débiles, es negado por un relato como el de Sartre y, en general, por aquellos discursos que terminan por mostrar un agente que a la hora de elegir carece de todo compromiso valorativo previo, entonces, se terminaría por pagar el precio de colocar al agente modélico que parece obtenerse de la pintura del pensador francés en una posición bastante cercana a la del simple sopesador de alternativas que se postula, según Taylor,

¹⁹⁹ “No podríamos *decidir* simplemente que la acción más significativa consiste en chapotear con los pies en barro tibio. Sin una explicación especial, no se trataría de una pretensión inteligible . . . no sabríamos qué sentido atribuir a alguien que supuestamente *pensara* que esto es así. ¿Qué podría querer *dar a entender* alguien que dijera esto? Pero si esto tiene sentido sólo después de una explicación (quizá sea el barro el elemento del espíritu del mundo con el que se entra en contacto gracias a los pies), queda abierto a la *crítica*. ¿Qué sucede si la explicación es falsa, si no tiene éxito, o puede ser sustituida por una descripción más apropiada? El que tengamos una cierta impresión de las cosas, nunca puede constituir base suficiente para respetar nuestra posición, porque nuestra impresión no puede *determinar* lo que es significativo. El relativismo blando se autodestruye. Las cosas adquieren importancia contra un fondo de inteligibilidad [. . .] una de las cosas que no podemos hacer, si tenemos que definirnos significativamente es suprimir o negar los horizontes contra los que las cosas adquieren significación para nosotros. Este es el tipo de paso contraproducente que se da con frecuencia en nuestra civilización subjetivista” (1994, p.72; las cursivas son de Taylor).

en los discursos utilitaristas. Ello tal vez muy a pesar de Sartre, dada su intención de mostrar la gravedad que reviste un dilema moral, así como la *incomensurabilidad* de las alternativas de elección que se le plantean al agente,²⁰⁰ características con las que también se encuentra de acuerdo Taylor, al ver en ellas aspectos fundamentales de su propia idea de una deliberación moral y del tipo de agente que la llevaría a cabo: su evaluador fuerte. De este modo, el autor canadiense ve en Sartre una suerte de inconsistencia que, a mi entender, creo que resulta bastante parecida a aquella que se señaló en Frankfurt: éste presenta un agente de carácter reflexivo, que se preocupa por ciertos asuntos, que evalúa sus deseos de acuerdo a sus preocupaciones fundamentales y que se identifica con algunos de sus deseos y, en cambio, se distancia de otros, hasta el punto de considerar impensable ciertas opciones de elección, al tiempo que, paradójicamente, se trata de un ser incapaz de deliberar, de ofrecer razones, e incluso de explicar/justificar sus elecciones.

En Frankfurt, pues, creo que acaba por producirse un resultado paradójico bastante parecido al que Taylor advierte en Sartre. A pesar de las enormes distancias que podrían señalarse entre una postura como la del filósofo francés y la que ejemplifican los pensadores utilitaristas, en el primero la figura del agente, que en principio anunciaba los mencionados aspectos de seriedad, reflexión y profundidad, termina por incluir características cercanas a las del evaluador débil que Taylor atribuye al modelo utilitarista. La razón de esta paradoja, y del resultado anticognitivistista al que ésta nos arroja, una vez más y como creo que sucede en Frankfurt, pienso que estaría en la negación de todo vínculo del agente con un horizonte valorativo previo a sus elecciones o, mejor, la concepción del *agente como un ser desvinculado* y carente de todo contexto referencial al que pueda acudir para deliberar y explicar sus decisiones. Esta figura del individuo solitario, autoconstituído y autosuficiente que decide en el vacío, que no conversa con otros agentes, que carece de contexto y de compromisos previos logra, pues, que se produzca un extraño encuentro entre una postura

²⁰⁰ Creo que este parentesco que cree encontrar Taylor, entre el agente que elige radicalmente y el sopesador de alternativas, tampoco sería del agrado de los utilitaristas, interesados en sostener un cierto cognitivismo ético —opuesto al irracionalismo que sostiene Sartre—, al tiempo que, para su idea de la deliberación moral requerirían afirmar, al contrario del autor francés, la conmensurabilidad de aquello que se juega en las opciones de elección; lo cual, como ya se ha comentado, facilita el cálculo maximizador que Taylor tanto rechaza.

anticognitivista como la de Sartre, con una posición de intenciones cognitivistas como la utilitaria.

Sartre quiere conservar la gravedad propia de los dilemas morales y, en este sentido, a los ojos Taylor, en la figura del agente que decide “radicalmente” se hallarían, repito, elementos fundamentales de su propia pintura del evaluador fuerte, pero la insistencia de Sartre en tornar al agente en un “inventor” de lo valioso, en un ser sin vínculos previos, le termina por llevar, según el autor canadiense, muy cerca de la figura del evaluador débil que elige, finalmente “porque sí”, “porque así me apetece”, o del simple sopesador del utilitarismo, que elige sobre la base de un cálculo de bienestar; lo cual seguramente no sería lo que el Sartre tendría en mente. Esto, pues, como he dicho, pienso que nos llevaría una vez más de vuelta a Frankfurt y a algunos de sus resultados aporéticos, concretamente, a su tesis de que los agentes humanos son seres para los cuales las cosas importan, seres que incluso son capaces de contravenir sus deseos por mor de aquello por lo que están fundamentalmente preocupados, pero que, finalmente, no podrían señalar razones para una toma de decisión como aquella que se ejemplifica en casos como el del chico del relato de Sartre, o el de Lutero, o el de la madre que se siente incapaz de entregar a su niño. Pues, recordemos, según la lectura que hace Frankfurt de estos ejemplos, los agentes terminan por decidir a causa de una necesidad volitiva que no es en absoluto discursiva y que les lleva a elegir aquella opción que acabe “pesando más” según lo determinado por la “fuerza” de dicha necesidad que, a la postre, resulta ser inarticulable; lo cual, como puede verse, parecería acercarnos a la idea sartreana de la elección radical.

4. A modo de conclusión: los límites de la autoaclaración y la esperanza en lo dialógico

Por lo anterior, pienso que parecería que tratándose de ambos pensadores, Sartre y Frankfurt, llegaríamos a una imagen que, desde una mirada como la de Taylor, resultaría bastante perturbadora: la de un agente que enmudece, que queda incapacitado para dar razones, lo cual, según el autor canadiense no resultaría una opción plausible, si bien, y como ya se ha dicho, de alguna manera Taylor sí cree que habría situaciones que permitirían que se hablara de una elección “radical”, pero se trataría de un uso especial de este adjetivo. Recordemos que sería una elección radical en el sentido de que las

alternativas de elección resultan inconmensurables y la lucha en la que se enfrasca el agente se debe justamente a su intento por alcanzar alguna forma de conmensurabilidad que, de no lograrse, sí causaría que entre dichas alternativas se tuviera que elegir radicalmente, al señalar modos de vivir completamente distintos, rutas que podrían señalar trayectorias vitales excluyentes, y de allí la angustia experimentada por quien busca autoaclararse y teme errar en dicho proceso.

With strong evaluation, [however], there can be and often is a plurality of ways of envisaging my predicament, and choice may be not just between what is clearly the higher and the lower, but between incommensurable ways of looking at this choice [. . .] we have here “objects” –courses of action– which can only be characterize through the quality of life they represent, and characterize contrastively. It is part of the desirability characterization of each that it has and undesirability story to tell about the other. But the struggle here is between [two] such characterizations, and this introduces a new incommensurability (1985c, pp. 26-27).

Más adelante Taylor afirma:

The real force of the theory of radical choice come from the sense that there are different moral perspectives, that there is a plurality of moral visions [. . .] (p. 33).

A mi entender, para Taylor, si bien la mencionada inconmensurabilidad justifica el uso del término “radical”, de todas formas la deliberación que permite articular las opciones en disputa autorizaría, si tomamos en serio la postura del canadiense, el que se insista en el carácter lingüístico y dialógico de dicho proceso de reflexión y deliberación, por oposición a la inarticulación y al solipsismo. En este punto creo que habría que señalar que, no obstante lo anterior, y aunque Taylor no lo señale expresamente, este papel del lenguaje y del dar razones pienso que *no* nos garantizaría que no haya errores, ni que la salida ante un dilema esté asegurada, ni tampoco que dicha salida termine en un final feliz. De allí que de nuevo me permita insistir en lo que arriba se comentó, a propósito de la “contaminación” y falta de autosuficiencia del agente: no se podría garantizar la transparencia de su proceso de autoaclaración, ni de aquello que en él es objeto de examen. Por lo cual, y por más que en Taylor haya una suerte de confianza en que se alcance una cierta claridad, de todas formas se me ocurre que al lector habría que anunciarle con un poco más de energía que el éxito rotundo no

está asegurado, ni en cuanto al logro de la claridad que busca alcanzar quien evalúa sus deseos y sus opciones de elección, ni en cuanto al dar razones de cara a otros agentes.

De allí que, si bien para nuestro autor sería necesario apartarse del irracionalismo, también creo que haría falta advertir sobre otro tipo de amenazas provenientes de otras formas de irracionalidad, a las que tal vez podríamos considerar de carácter más “estructural”, y sobre las que, a mi entender, tal vez a Taylor le convendría hacer un mayor énfasis. Me refiero no sólo a los peligros relacionados con las dificultades inherentes al proceso de autoaclaración, o con las posibilidades del autoengaño, las cuales, a mi entender, siempre estarán presentes como tales peligros si partimos de un sano falibilismo, sino también con las maneras en que el agente pueda internalizar las voces de esos otros en cuya lengua y bajo cuyas reglas debe articular las razones para su acción, v.g. las posibilidades de manipulación y del uso del poder que bien puede estar presente en nuestro conversar con nuestros “otros”, ya sea *in foro interno*, ya sea públicamente. Esto creo que nos tendría que poner sobre alerta, como también creo que Taylor de alguna manera lo advierte, sobre las implicaciones del hecho de que la articulación de aquello que resulta más valioso para nosotros ocurra mediante un proceso de carácter dialógico, en el cual —usando el lenguaje que hemos aprendido de otros— estemos en posición de conversar con esos otros.

Pienso que esto tendría que hacernos caer en cuenta del aspecto fáctico, histórico de este proceso. Nuestra historia como individuos y como parte de un grupo o de una cultura, historia en la que hemos forjado y negociado nuestra identidad personal/colectiva, y en la cual hemos gozado del reconocimiento que nos hayan otorgado o, por el contrario, hayamos sufrido por la falta de tal reconocimiento, así como nuestra historia como hablantes de una comunidad para la cual ciertos valores son fundamentales, creo que representan una carga fáctica de la cual tendríamos que hacernos concientes, dado lo determinante que ésta pueda ser para el proceso de aclaración de nuestra propia identidad. Lo cual, a mi entender, nos haría difícil afirmar que el horizonte valorativo de cada quien, horizonte a partir del cual se evalúan los propios deseos y preferencias, pueda ser igualmente accesible para todos, con lo cual ser un evaluador fuerte no sería una posibilidad igual de accesible para todos. Pienso que tampoco podría convenirse en que dicho horizonte venga determinado por el referente normativo, universalmente compartible de la “naturaleza humana”. Más bien, creo que

podríamos decir, con Taylor, que las versiones acerca de lo valioso, en el caso de los dilemas morales, nos muestran que lo que está en juego son las diferentes formas posibles de llevar vidas humanas. En ello, la identidad que se busca desvelar mediante el proceso de aclaración de las propias y más fundamentales evaluaciones, no es la de “cualquier” agente, sino la de cada uno. Pero, insisto, creo que esto nos tendría que hacer ver que la situación de cada uno es distinta en muchos aspectos, incluyendo las posibilidades mismas que cada quien tenga, dada su especial situación en su contexto social, de ser reflexivo, conciente y libre de autoengaño o de manipulación al momento de intentar autoaclararse. La dificultad para la transparencia explica lo que bien señala Taylor: la angustia que experimenta quien se enfrenta ante un dilema moral en buena parte se debería a que la persona intentaría aclarar unos compromisos que no están determinados de antemano por una instancia igual de transparente y accesible para todos. No sería, pues, un referente universal e impersonal que le ahorraría al agente el esfuerzo de identificar sus más fundamentales evaluaciones, sino que le obligaría a una labor signada por la falibilidad: la labor de autodescubrirse. Pero, insisto, a esta dificultad habría que agregar aquella que acaba de mencionarse: esta autoaclaración que el agente debe llevar a cabo, tiene lugar en una situación “suya”, única, irreducible a la de otros y, sobre todo, signada por los peligros que conlleva la lucha por el reconocimiento de esos otros.

Sin embargo, creo que si aunáramos el aspecto dialógico del proceso de autoaclaración al hecho de que el horizonte de valor y de sentido es precisamente eso, un horizonte, un trasfondo cuya vía de descubrimiento y aclaración no tiene por qué tener un punto final, pienso que el carácter falible del agente y de los resultados siempre parciales a los que llegue, admitirían ser vistos de dos maneras: no sólo como una fuente de peligro como la que acaba de mencionarse, sino también como aquello que nos haría sostener una suerte de confianza en lo dialógico y en el dar razones. Pues el proceso mediante el cual nos autoaclaramos, conversando con nosotros mismos, conversando con otros, unas veces más y otras veces menos pacífica y/o duramente, siempre admitiría revisiones. Siempre cabría la posibilidad de que cada nuevo intento de articular nuestras evaluaciones más fundamentales altere nuestra experiencia y cambie nuestra sensibilidad, de tal manera que podamos cambiar nuestra versión acerca de qué

es lo más valioso, y de quiénes somos nosotros mismos.²⁰¹ En palabras de Taylor, toda evaluación puede admitir una re-evaluación (1985, p. 40), y esto significa que obligamos a lo inarticulado, aquel *topos* en el que parecíamos hallarnos ante una aporía como la que cabría verse en Frankfurt, a desvelarse mediante el lenguaje y ser parte del dar razones; razones que pueden ser siempre objeto de revisión, sobre todo cuando nos encontramos ante la presencia de otros agentes y sometidos a los avatares de nuevas experiencias sobre las cuales conversar –repito, unas veces más, otras veces menos pacífica o duramente– con ellos.

Si bien la relación con esos “otros”, en mi opinión, es asumida por Taylor bajo una luz que acaso pueda ser criticada por cierto exceso de optimismo,²⁰² también pienso que, desde el punto de vista de este autor, el peso que dicha relación tenga sobre nosotros no siempre impediría el ejercicio de nuestra capacidad crítica. Por lo tanto, pienso que, con Taylor, no estaríamos condenados a asumir una figura que aparecía en el discurso de Frankfurt y que creo también resultaba un tanto perturbadora, e incluso poco sana y poco realista: la de un agente que pareciera estar impedido para el cambio, para la evolución y, con ello, para el diálogo, la crítica y la discusión con otros, al estar atado a la “necesidad” de sus voliciones de orden superior. Esto es, a la “necesidad volitiva” con la que se le planteaban sus preocupaciones fundamentales y, por tanto, a una suerte de identidad inamovible que terminaba lastrándole con un blindaje ante la crítica. En este sentido, a pesar incluso de lo que el propio Frankfurt tal vez quisiera, pienso que su modelo de agente en algunos aspectos podría recordarnos algunas de las características menos deseables del maximizador autointeresado y de sus “preferencias”.

De allí que crea que, si bien cabría la posibilidad de hacer una cierta lectura de Taylor según la cual su noción de “identidad” podría ser proclive a padecer los peligros de una merma en la autonomía del agente, sin embargo es posible mostrar que, si nos situamos en un punto de vista como el suyo, podríamos estar más a salvo de la idea de una identidad inamovible tal y como la que resultaría del compromiso de Frankfurt con su noción de necesidad volitiva, más exactamente, con la idea de unos “límites” que, según este autor, el agente no podría traspasar sin pagar el precio de “estallar” como tal agente, así como la conexa noción de “ambivalencia”, entendida como una suerte de

²⁰¹ Al respecto véase: Thiebaut (2008).

²⁰² Cf. Thiebaut (1994 y 2009).

infortunio insuperable y destructivo para la propia capacidad de agencia. Por ende, también pienso que cabe encontrar en Taylor una vía para estar más a salvo del horror de Frankfurt ante esta experiencia de la ambivalencia que, según él, viviría el agente que tuviera dificultades para elegir entre dos de sus voliciones de orden superior. Pues el horizonte de los compromisos valorativos, tal como es presentado por Taylor, sería para el agente un trasfondo que en algunas ocasiones, como cuando se está ante un auténtico dilema moral, podría tornarse en problemático debiendo, por tanto, ser objeto de aclaración.

De manera que ni la naturaleza de dicho horizonte, ni el carácter de la mencionada aclaración autorizarían a temer como indeseables de suyo, o como auténticas catástrofes, las posibles faltas de claridad, los procesos en los que se experimenten dudas y ambivalencias, como si éstos fuesen necesariamente los causantes de una eventual y aterradora quiebra en la capacidad de agencia de una persona. Antes bien, creo que si asumimos el esquema de Taylor, dichos procesos serían parte de lo que no podemos evitar al ser los agentes humanos que somos. Incluso creo que hasta podrían ser pensados como oportunidades de ejercitar, de tanto en tanto, una sana labor de profilaxis, de reflexión y de autocrítica que, si bien en algunas ocasiones, como se ha visto, podrían conllevar dolor y lucha, no tienen que ser valoradas como esencialmente destructivas, ni como el objeto de horror con que Frankfurt parece identificarles. Así, puede que (dicho en una suerte de jerga husserliana) cotidianamente operemos desde la “actitud natural” y, por ende, que lo usual sea que normalmente supongamos ese horizonte de sentido y de valor y, por tanto, que sólo en algunas ocasiones nos veamos impelidos a examinarlo. Empero, son dichas ocasiones de crisis y de duda las que justamente nos harían caer en la cuenta de que, entonces, aquello que se daba por sentido ya no está tan claro, ni tan libre de dudas, ni tan inmune a la crítica, y que conviene por tanto comenzar a hablar de ello, e intentar explicitar eso que ahora se torna en un problema; entre otras razones, porque también puede ocurrir que haya momentos en los que convendría que sea experimentado como un tal problema.

Por supuesto, en este punto alguien podría señalar que con estas afirmaciones y expresiones de entusiasmo, se podría estar pecando de un cierto exceso de optimismo en favor de una postura cognitivista, optimismo que sería análogo al que arriba se menciona como atribuible a Taylor, en lo que se refiere a su idea acerca de nuestra

posibilidad de no quedar aplastados bajo el peso que puedan tener nuestros “otros” y nuestro contexto cultural, en detrimento de nuestra capacidad crítica. Pero creo que, si bien es cierto que cabe la posibilidad de que alguien se niegue, en algunas ocasiones, a ofrecer tales razones, o incluso, si alguien se encontrara, bien sea oprimido, bien sea cegado por el mencionado peso de su contexto, también pienso que en Taylor pueden hallarse algunas bases para una confianza no ingenua en que podamos considerar que este caso del sujeto que, o bien no logra tomar la distancia crítica que le permita “verse” a sí mismo y a su situación y, por ende, no logra argumentar, o bien se sienta tan cómodo en dicha situación que sencillamente no quiera hacerlo, aunque se trate de un caso ilustrativo de lo que muchas veces hacemos o nos sucede, sin embargo no tendría por qué ser asumido como “inescapable”.

Si esto último fuera así, es decir si en los aportes de Taylor se pudieran hallar tales bases, entonces pienso que tampoco tendríamos por qué considerar que dicha postura de inaccesibilidad o insensibilidad a las razones, a la crítica y a la reflexión, por más frecuente que sea, haya de tener ni un peso normativo ni la carga de una fatalidad insuperable. De allí que, si podemos convenir con Frankfurt en que los agentes humanos tenemos la posibilidad de ser reflexivos y evaluar nuestros deseos, y no sólo la posibilidad de hacer cálculos orientados a la satisfacción de éstos, entonces también creo que, a pesar de las reticencias de este autor frente a ese “dar razones”, tanto su argumentación en pro la posibilidad de una tal evaluación, como su postura frente al papel determinante que le atribuye al *caring* de los agentes humanos, tendrían que ser incluidas dentro una argumentación que contribuya a sostener las mencionadas bases cognitivistas –ofrecidas, a mi entender, por Taylor– sobre las cuales sustentar la confianza en las posibilidades del “dar razones”. De lo contrario, dejaría de tener sentido la insistencia de Frankfurt en establecer una distinción entre quien asume la actitud de preocuparse por la calidad de sus preferencias y, por contra, la actitud de quien no se preocupa de tal modo y, por tanto, se comporta como un *wanton*. Pienso que efectivamente bien podemos comportarnos de esta segunda manera, pero esto no significa que este hecho (que a veces podría incluso obedecer a una sana necesidad), tenga que ser visto, repito, como poseyendo de suyo un peso normativo frente a la otra postura, la del agente evaluador (y autoevaluador) y, ni siquiera, como teniendo el peso de lo inevitable.

Por lo anterior creo que, de manera similar, es cierto que en muchas ocasiones podemos hacer uso de la razón como si fuera –parafraseando la máxima de Hume– una “esclava de las pasiones”, pero también creo que esto no implica que en otras ocasiones no tengamos ante nosotros la oportunidad, e incluso la necesidad y la obligación de distanciarnos críticamente de dichas pasiones, y utilizar nuestras capacidades deliberativas para otros menesteres; entre otros motivos porque, como bien le señala Scanlon (2002) a Frankfurt, tampoco sería muy acertado partir de una escisión tan radical entre “razones” y “pasiones” o (si se me permite decirlo en estos términos) entre la razón y su “otro”, como si dicha escisión fuera evidente de suyo, como si no se pudiera sospechar de ella, por lo menos teniendo en cuenta que la persona es “una”, a pesar de las diversas facetas que pueda (y, repito, a veces sea sano que “deba” presentar). Si fuera plausible tomar tal distancia crítica, creo entonces que, con Taylor, sí podríamos al menos intentar “ir más allá de Hume” y, con ello, tratar de ofrecer una idea más incluyente de agencia, de motivaciones y de razonabilidad humanas.

Conclusiones

In assessing the opportunity aspects of freedom, the focus has to be on the alternatives that a person has reason to value or want . . . thus an assessment of the opportunities a person has would require some understanding of what the person would want to have and have reason to value having . . . freedom can not be fully appraised without some idea of what a person prefers and has reason to prefer. Thus, there is a basic use of rational assessment in appraising freedom, and in this sense, freedom must depend on reasoned assessment of having different options . . . Rationality as the use of reasoned scrutiny cannot but be central to the idea and the assessment of freedom . . . the concept of rationality must accommodate the diversity of reasons that may sensibly motivate choice.

(A. Sen)

Con la elaboración de este trabajo se buscaba llevar a cabo una indagación sobre las posibilidades atribuibles a una forma de agencia humana a la que se pudiera pensar como capaz de superar los límites impuestos por el modelo del *homo oeconomicus*. Las intuiciones sobre dichas posibilidades en buena parte se fueron aclarando a lo largo del proceso de elaboración del presente trabajo, en una primera instancia, gracias a la revisión de las características que se descubrieron en el agente modélico propuesto por Gauthier y, luego, en un segundo momento, en los aportes que se hallaron en los discursos de Frankfurt y de Taylor. A lo largo de los tres primeros capítulos, se examinaron las implicaciones del compromiso inicial que asume Gauthier con una idea exclusivamente maximizadora del agente y de sus capacidades deliberativas, implicaciones que incluían el privilegiar al mercado como modelo de interacción social, al agente económico como modelo de agencia, no sólo racional, sino también moral, y al autointerés como el motivador y el fundamento de la moral. En este sentido, dicho fundamento también proveería la base de legitimidad de los acuerdos políticos, en especial, del contrato social.

Lo anterior condujo a ciertos problemas relacionados con las tensiones entre la agencia económica y la agencia moral, sobre todo si se tiene en cuenta el sesgo que podría traer el egoísmo atribuible a quienes transan en el mercado y, por ende, la posible falta de estabilidad y de fuerza vinculante que cabría esperar de unos acuerdos suscritos por este tipo de agentes, a los que se dibuja bajo la figura de un Robinson que quisiera verse libre de todo tipo de ataduras que le signifiquen una merma en su libertad,

entendida como la ausencia de trabas a la conducta maximizadora. Se examinó la estrategia de Gauthier para responder a estos problemas, en primer término, mediante su teoría de la justicia, la cual, en contraste con la de J. Rawls, asumía como canónico el modelo de racionalidad desarrollado por la teoría de la elección racional, así como el modelo de un agente que, a diferencia del “insensato” frente al que argumenta Hobbes, está dispuesto a cumplir con el pacto social, pero que también, en contraste con los agentes de la posición original de Rawls, se trata de un agente maximizador racional preocupado por su propio beneficio, razón por la cual, a los ojos de Gauthier, sería un agente ideal-imparcial con el que cualquiera de nosotros se podría identificar.

En segundo lugar, las dudas que pudieran surgir acerca de la moralidad de los acuerdos a los que puedan llegar unos agentes maximizadores dispuestos a cumplir sus pactos únicamente en razón del beneficio que calculen pueden obtener de la cooperación social, son enfrentadas por Gauthier mediante su estrategia de “pacificar”, tanto a este tipo de agentes, para que resulten menos proclives a la insociabilidad, como al escenario en el que tales agentes se interrelacionan: el mercado. De este modo, se examinó la forma en la que, hacia el final de su argumentación en *La moral por acuerdo*, Gauthier introduce ciertos matices y correcciones con los que intentaría conjurar este tipo de peligros debidos al egoísmo y a la falta de compromiso moral atribuible al *homo oeconomicus*. Dichos matices y correcciones le permitirían al autor alejarse un poco de esta figura, o mejor, hacerle ciertos cambios orientados a convertirla en el “individuo liberal”. En contraste con el *homo oeconomicus*, en este último, según Gauthier, se destacarían, entre otros aspectos, su mayor sociabilidad y su mayor autonomía.

No obstante, también se vio que dicha sociabilidad obedecería a unas motivaciones –las motivaciones “secundarias”– a las que el autor considera como superpuestas a las auténticas motivaciones –las motivaciones “primarias”– que son aquellas que, para Gauthier, realmente explican la conducta de los agentes, la cual sigue siendo, en su origen y en lo fundamental, maximizadora y egoísta. Así, aun a pesar de que los propios agentes puedan pensar que tienen un interés *no* instrumental en sus congéneres y en las relaciones que con ellos establezcan, para Gauthier dicho interés obedecería, en último término, a razones de tipo evolucionario, esto es, a la utilidad adaptativa que han tenido las relaciones sociales para el beneficio de los agentes

individuales. Así, el autor intenta justificar su compromiso con el ideal libertario del individualismo mediante su exposición de la figura del agente liberal, de tal modo que se muestre a la maximización de la utilidad individual como la base de aquello que “en la superficie” podría aparecer como un gusto por la sociabilidad o un interés no instrumental, desarrollado por este tipo agente, con relación a sus congéneres.

Se vio, entonces, que en Gauthier puede advertirse un cierto compromiso reduccionista, una de cuyas caras podría mostrarse en la explicación que el autor ofrece sobre este aspecto “sociable” del individuo liberal, explicación según la cual dicha sociabilidad obedece a una necesidad adaptativa, de la cual no son concientes los agentes que tomen parte en la praxis moral, pero que sí es accesible para quien conoce la “verdadera” explicación: el filósofo moral que, desde “fuera”, “observa” y “objetiva” las interacciones en las que se involucran los agentes así “observados”. Se señaló, entonces, la relevancia de esta actitud asumida por Gauthier, al marcar un contraste con los discursos de Frankfurt y de Taylor, quienes sí asumirían el punto de vista de los agentes que toman parte en la vida moral.

De otra parte, como arriba se menciona, Gauthier también intenta enfrentar los peligros de una agencia meramente autointeresada, mediante la exposición de otro rasgo importante de su agente liberal, el cual, en contraste con el maximizador egoísta, podría ser pensado como teniendo una mayor autonomía. Ésta se vería expresada en lo que el autor considera es una muestra de un mayor poder autoevaluativo: la preocupación que tendría el individuo liberal por sus preferencias futuras. Con ello, este personaje también estaría más cerca (de lo que estaba el *homo oeconomicus*) de tener una identidad y una unidad que le permitiría persistir como el mismo agente a lo largo del tiempo, sin el peligro de una posible dispersión ocasionada por la diversidad y falta de conexión que podrían presentar las preferencias que vaya teniendo en el transcurso de su vida.

No obstante, al examinarse los aportes de Frankfurt y, sobre todo, los de Taylor, en los dos últimos capítulos, se hizo más claro que, al lanzar una segunda mirada sobre el discurso de Gauthier, esta noción de “vida” difícilmente tendría cabida dentro de dicho discurso, dado que el maximizador egoísta de Gauthier sigue siendo un ser abstracto, sin vínculos que le aten a sus semejantes y, por tanto, sin una historia compartida con ellos. De este modo, en la segunda parte de la tesis, al analizarse la propuesta de Frankfurt, se vio que, aunque este autor comparte con Gauthier la idea de

un agente solitario, sin vínculos con un contexto social y, por tanto, pensado en clave monológica, sin embargo, en el agente dibujado por Frankfurt –la “persona”– podían advertirse elementos de una mayor complejidad que significarían un importante salto cualitativo frente al agente de Gauthier, elementos que el propio Frankfurt resume en el hecho de que su agente modélico tendría una “vida integrada”, en contraste con el tipo de agencia criticado por este autor: aquella a la que se refiere como “fragmentaria”, al ser propia de un sujeto cuyos deseos son una suerte de episodios que meramente “le ocurren” y por los cuales el agente no asume ninguna responsabilidad.

Antes de recordar los aportes de Frankfurt, quisiera detenerme en un elemento, a mi entender, clave en la figura del agente, tal y como lo entiende Gauthier, bien sea en su versión inicial, la del *homo oeconomicus*, bien en su versión final, la del individuo liberal. Creo que buena parte del atractivo que tienen ambas figuras dibujadas por Gauthier, incluso si por ello se tiene que pagar el precio de presentarlas de un modo artificioso, al verlas como seres sin contexto, sin vínculos sociales, sin una historia compartida con otros, sin un lenguaje en el cual entenderse con esos otros o, mejor, sin un lenguaje aprendido de esos otros, sin la influencia que éstos hayan tenido sobre la persona, influencia que incluso puede moldear la propia autoimagen de la persona, es un atractivo que reside precisamente en la posibilidad de liberarse de dicha influencia.

La idea de un ser que no depende de otros, ni siquiera para pensarse a sí mismo o mejor, menos aun para pensarse a sí mismo; un ser que antecede al mundo social y lo moldea dependiendo de sus intereses y, por tanto, se vea libre de las ataduras que tendría un sujeto que sea moldeado por un contexto social que le anteceda y al que se vea “arrojado” antes de que pueda hacerse cargo de dicho contexto y cómo éste pueda afectarle; la idea de un agente que “se inventa” el mundo a la medida de sus deseos y que goza de plena capacidad de agencia antes de asociarse –libremente, por supuesto– con otros agentes, creo que expresa un atractivo ideal de libertad, aunque con las limitaciones que impone la figura de este tipo de agente y con las formas de deliberación que se le atribuyan, como veremos más adelante, paradójicamente se propicie todo lo contrario: una extraña falta de libertad.

Sobre esto último se volverá al final de este texto, pero antes de llegar allá, habría que mencionar que este ideal expresado en un agente como el que dibuja Gauthier, cuenta además con la ventaja de que va acompañado del modelo de una

racionalidad exclusivamente maximizadora, tal y como ésta es desarrollada por la teoría de la elección racional, lo cual significa contar con un soporte que lo muestra como un ideal no sólo posible –lejos de los artificios arriba mencionados– sino también “riguroso”. Pues, finalmente, dicha forma de racionalidad puede ser defendida como “la” racionalidad, dado que no podría ser puesta en cuestión, al tener que ser incluida en cualquier relato sobre las formas posibles de deliberación. Pienso que esto muestra la pertinencia de examinar una propuesta como la de Gauthier, y de allí también el atractivo que puede tener su figura del individuo liberal, en el que se ven matizados algunos de los rasgos menos alentadores del *homo oeconomicus*, matices que incluirían la posibilidad de que el liberal se preocupe por sus preferencias futuras.

Una vez examinadas las consecuencias de tales matices introducidos por Gauthier, se procedió a analizar aquellas características del agente modélico propuesto por Frankfurt, características en las que podría verse un salto cualitativo importante con respecto del agente dibujado por Gauthier. Dicho salto pudo apreciarse, de una parte, en el hecho de que el agente de Frankfurt contaría con una base motivacional mucho más rica, representada por aquellos asuntos sobre los cuales el sujeto se encuentra fundamentalmente *preocupado*, más allá de consideraciones sobre su propio beneficio. En estrecha relación con la naturaleza de esta base motivacional proporcionada por las preocupaciones fundamentales de la persona, en Frankfurt encontramos, en segundo término, la novedad de que este tipo de agente es, precisamente, aquel que es capaz de *evaluar* sus deseos, preocupándose por la deseabilidad o indeseabilidad de éstos, identificándose con algunos de ellos y, por el contrario, rechazando otros. Se trata, pues, de un sujeto que se preocupa por su identidad y que trata de conservarla a través del tiempo, asumiendo frente a sus deseos una actitud crítica que revela una mayor profundidad y una capacidad evaluativa y autoevaluativa que le permitiría superar algunas de las limitaciones más importantes que presentaba el maximizador egoísta de Gauthier. Así mismo, dicha capacidad estaría estrechamente relacionada con una mayor riqueza y complejidad atribuible a los deseos y motivaciones que tal agente manifiesta, con lo cual se habría dado un paso importante más allá de la opacidad y simplicidad que esconde o suprime dicha riqueza bajo la escueta fórmula del “autointerés”.

En este punto, pienso que conviene anotar, a la luz de una revisión más “global” de los tres modelos de agencia que se estudiaron, que habría en Frankfurt un aspecto

destacado de su análisis de aquellos elementos a los que se refiere como “complejidad volitiva”, “necesidad volitiva”, e “identificación incondicional”: el hecho de que tales elementos de la agencia sean concebidos como “logros”, como producto de un esfuerzo, de una lucha mediante la cual el agente intenta “ganarse para sí mismo”, unificarse en su voluntad, en síntesis, lograr ser propiamente un agente. Este aspecto del análisis de Frankfurt bien podría justificar un llamado de atención, no sólo para quienes asuman el modelo de agencia supuesto por discursos que, como el Gauthier, muestran un agente que aparece en el escenario social con plena capacidad de agencia, siendo por completo dueño de sí mismo, tanto en lo que respecta a su conciencia de sí, como a su poder para llevar a cabo sus elecciones; un agente con sus poderes deliberativos y su capacidad reflexiva completamente desarrollados, hasta el punto en que, como arriba se menciona, dicho escenario social se construye mediante la libre asociación de unos individuos pensados de este modo. En Frankfurt tal vez se cometa el error, también compartido por Gauthier, de no tener en cuenta la importancia que tendría tal escenario para las posibilidades del desarrollo y del ejercicio de la agencia individual, pero creo que ello posiblemente se debería a su preocupación por poner en un primer plano este aspecto de la lucha por ganarse a sí mismo. Esta lucha se vería de manera más explícita en Taylor, como en breve se comentará, sin necesidad de obviar la relación con los otros, sobre todo si se tiene en cuenta su análisis de los dilemas morales y lo que, con motivo de éstos, se desarrolla como un esfuerzo del agente por aclararse a sí mismo.

No obstante los significativos avances que, frente al agente modélico de Gauthier, representa la figura de la “persona” de Frankfurt, sobre todo si se tiene en cuenta su idea de que este tipo de agente es un ser capaz de evaluar sus deseos, al estar preocupado por la deseabilidad de éstos, y al otorgarle para ello un peso especial a sus “preocupaciones” (*caring*) fundamentales, al final de la exposición de la propuesta de Frankfurt nos encontramos con algunos resultados aporéticos que se atribuyeron a su compromiso con una postura anticognitivist, en razón de la cual su agente modélico carecería de referentes normativos que le permitieran juzgar dicha deseabilidad/indeseabilidad de sus deseos.

Esto le llevaba, entre otros resultados problemáticos, a no poder justificar la propia evaluación de los deseos y caer en una suerte de decisionismo que pondría en peligro las conquistas alcanzadas frente al maximizador egoísta de Gauthier. Ante tales

resultados, se propuso, como una salida que permitiera salvar las mencionadas conquistas, una lectura de los metadeseos de Frankfurt en virtud de la cual éstos pudieran verse como haciendo parte de un horizonte de sentido, en el cual los referentes a los que apele el agente para evaluar sus deseos no sean vistos como conformando una escala jerárquica, marcada por la relación fundamento-fundamentado, sino como los hilos de una compleja “red” de significados.

Se hizo énfasis en esta figura de la “red”, con el fin de que se pudiera entender mejor la naturaleza de la “lectura” que el agente puede hacer de sus motivaciones y preocupaciones. De tal manera que éstas se inserten en un complejo holista de sentido y, con ello, se evite el peligro de una cadena infinita de deseos-metadeseos que, al no poder descansar sobre un último “fundamento”, se viera signada por el absurdo. Esta idea de una “red” de referentes de sentido se hizo más clara en la propuesta de Taylor, gracias a su tesis sobre el papel que cumple, en los juicios de los agentes, aquello que este autor concibe como un horizonte de valor y de sentido sin el cual dichos juicios no podrían tener lugar. Se propuso, entonces, una lectura de la propuesta de Frankfurt a partir de las tesis de Taylor, de tal forma que, desde este último, se pudieran entender, desarrollar y radicalizar los aportes del primero. Así, la evaluación de los deseos, tal y como es concebida por Frankfurt, podría ser pensada como expresiva de una forma de agencia más compleja e incluyente, de la cual no tendrían por qué quedar expulsados los elementos cognitivos, ni la apelación a referentes que el sujeto pudiera compartir con otros agentes.

En Taylor, dicha apelación se lleva a cabo sobre todo en aquello que este autor entiende por “evaluaciones fuertes”, en tanto que en ellas el agente evalúa sus opciones de elección a la luz de unos referentes de *valor*, los cuales, a su vez, se relacionan con las *formas de vida* de las que tales referentes son una expresión. Dichas formas de vida se constituirían en complejas unidades de sentido las cuales proveerían un marco para la reflexión del tipo de agente que lleva a cabo esa particular manera de evaluar las propias motivaciones: la evaluación fuerte. Taylor contrasta esa forma de evaluación con aquella otra a la que se refiere como “evaluación débil”, y que el autor atribuye de manera especial al agente modélico que, según él, se privilegia en el discurso utilitarista: el “simple sopesador” de alternativas.

Un agente pensado de este modo, para Taylor, se limitaría a un cálculo estratégico que le permitiese establecer prioridades entre sus deseos, sin que en ello juegue ningún papel la cuestión del valor, ni la identificación con una forma de vida, que es, por el contrario, lo que distingue aquello que trata de elucidar el evaluador fuerte mediante su proceso de reflexión, proceso en el cual la dificultad estriba en llegar a una autointerpretación que resulte la menos engañosa, la más reveladora de aquello que para el agente sea lo más valioso, o provea las fuentes de sus compromisos más importantes. En este sentido, pienso que, al igual que en Frankfurt, se pone de presente aquello que arriba se menciona como el aspecto de “lucha” que habría en este esfuerzo que lleva a cabo el agente por “ganarse a sí mismo” (o “para sí mismo”).

Se vio que una de las “armas” más importantes de las que tiene que valerse el agente de Taylor para poder librar esa “lucha”, y lograr ser un evaluador fuerte, es su lenguaje; concretamente, el uso de un lenguaje que le permita caracterizar sus opciones de elección mediante términos valorativamente cargados, que señalen contrastes fuertes, gracias a los cuales dichas opciones de elección adquieren su sentido como expresivas de una cierta forma de vida, y puedan ser articuladas al modo de “autointerpretaciones”, las cuales, a diferencia de lo que ocurriría con un relato sobre eventos externos al sujeto, cambiarían al propio agente. Éste no quedaría “igual” después de cada esfuerzo de autointerpretación, ni en cuanto a cómo “habla” de sí mismo y de sus experiencias como agente, ni en cuanto al tipo de experiencias que podría vivir y entender. Cambiarían, por tanto, su horizonte de comprensión, como su forma de actuar.

El uso de este tipo de lenguaje señalaría un distancia importante frente al lenguaje neutro que caracteriza al evaluador débil, lenguaje que estaría más cerca del proyecto reduccionista de traducir los términos cualitativos a términos observacionales, “purgados” de toda carga valorativa que impidiese que la deliberación moral se pudiera adelantar al modo de una operación de cálculo. Son estos contrastes valorativos fuertes los que permiten, por ejemplo, que un acto sea visto/juzgado como “cobarde” y se lo entienda como opuesto a otro acto al que se califica de “valiente”. Pienso que este tipo de lenguaje nos alejaría del agente de Gauthier, en tanto nos muestra de manera más clara, ahora bajo la luz que pueda arrojarse desde el discurso de Taylor, cómo es que el maximizador egoísta aparecería como un ser cuyas capacidades reflexivas estarían restringidas por el tipo de lenguaje que pueda usar —sólo para sí mismo, ya que no

parece conversar con otros— al momento de deliberar y decidir, pudiendo describir sus opciones de elección únicamente mediante un lenguaje que permita verlas como conmensurables, de modo que entre ellas las diferencias relevantes sólo sean de carácter cuantitativo.

De este modo, al examinar las implicaciones que tendría el uso de un lenguaje que, por el contrario, permita dar cuenta de las diferencias cualitativas, se vio como un resultado importante la posibilidad de que el evaluador fuerte, a diferencia del evaluador débil, pudiera adelantar su proceso de deliberación mediante una descripción más rica de cada una de sus opciones de elección, lo cual, a su vez, permitiría una decisión más calificada, para la cual pudiera ofrecer *razones*, las cuales también podrían ser razones articulables ante otros agentes. Por el contrario, el evaluador débil, al dar cuenta de sus decisiones y de sus opciones de elección llegaría, finalmente, a aquello que Taylor llama la “experiencia de la inarticulación”. Pues, al carecer de un lenguaje como el del evaluador fuerte, la decisión final del evaluador débil no podría ser explicada apelándose para ello a una rica descripción de sus opciones de elección, de modo que entre ellas no podría apreciarse propiamente lo que contaría como “razones” para tal elección.

Creo que en este aspecto, un discurso como el de Taylor podría apoyar una postura cognitivista desde la cual se logre mostrar que sí hay salidas, distintas tanto: 1) Al irracionalismo o al decisionismo, como a 2) Un cognitivismo atado a un modelo exclusivamente maximizador, tanto de la agencia humana, como de la racionalidad práctica. Con ello, pienso que en Taylor se hallarían recursos importantes para hacer contrapeso al innegable poder que ha tenido esta segunda forma de entender “lo racional”, poder que acaso esté estrechamente conectado con el rechazo que pudiera provocar la primera opción, el irracionalismo, rechazo del cual la propuesta Gauthier podría ser una expresión.

Esta posibilidad para el cognitivismo, que podríamos hallar en la forma en que, desde un discurso como el Taylor, se concibe a las “razones para actuar”, al proceso mediante el que puede arribarse a dichas razones, y al agente que reflexiona sobre éstas, se aprecia en la discusión que el autor plantea con relación al tema de los dilemas morales. Taylor critica lo que él considera es una forma errónea de entender, tanto la agencia humana, como la naturaleza de la deliberación moral, y aquello que hace que se

presenten auténticos dilemas morales. Este error, según Taylor, es cometido por Sartre en su concepción de dichos dilemas, al considerar que éstos serían decidibles mediante una “elección radical” para la cual el agente no podría, finalmente, ofrecer razones. Al contrario de lo que Taylor le atribuye a Sartre, el primero muestra que, si el agente que se enfrenta a un dilema moral lo experimenta como tal dilema, es porque cada una de las opciones en disputa se le presenta como un reclamo moral justificado y, por tanto, no serían opciones que éste “se invente” o “cree” en el vacío, sin ataduras previas con otros agentes, o con ciertas formas de ver y experimentar el mundo, sino que resultan significativas para el sujeto sobre la base de un horizonte de valor y de sentido que, por su misma naturaleza y por la especial forma de vinculación que el agente tiene con él, no ha sido tampoco un objeto de “invención” ni de “elección”. En este sentido, se advirtió en Taylor la cercanía con una idea ya analizada en la parte dedicada a Frankfurt: la curiosa mezcla de necesidad y libertad que marca la relación del agente con sus preocupaciones fundamentales, mezcla que incluso le hace ver al agente de Frankfurt que la contravención de éstas serían una suerte de traición a sí mismo, algo que sería “impensable” para la persona, en el sentido de transgredir los límites de aquello que el sujeto ve como los límites de lo que puede hacer, o mejor, de lo que puede “considerar” hacer.

Ahora bien, en Taylor, la relación del agente con ese horizonte valorativo que la persona supone en sus juicios y elecciones, relación que explica el modo en que le “reclaman” sus opciones de elección en el caso de un dilema moral, sin embargo no implica que dichas opciones o dicho horizonte estén completamente claros para el sujeto. Éste, por tanto, y con ocasión del dilema, se siente impelido a aclararlos, y por esta misma vía, aclarar eso que supone como lo más valioso, eso que hace parte de su horizonte de valor y de sentido. Para lograr dicha aclaración, la persona se esfuerza por describir cada una de sus opciones de elección de la manera más rica y fiel posible, buscando con esto arribar a una decisión para la que pueda hallar razones; razones en las cuales el agente se reconozca, pero que también serían razones que éste pueda articular de cara a otros agentes.

En este punto, conviene recordar que este resultado tiene la relevancia de que muestra que la naturaleza de los referentes a los que apela la persona en su intento por lograr dicha articulación rompería, entonces, con un esquema monológico como aquel

desde el cual se defiende la idea de una “elección radical”, pero también como el que se supondría, desde otra orilla, en discursos como el de Frankfurt, pero sobre todo, en el de Gauthier, si bien, por supuesto, entre estos dos autores se deben reconocer diferencias importantes, como las ya anotadas.

De allí que, a propósito de este asunto, se advirtiera cierto paralelismo entre las falencias que Taylor cree hallar en la idea de la “elección radical” de Sartre, y algunos de los problemas señalados en la parte dedicada a Frankfurt, concretamente, el hecho de que en este autor también se llegaría a la experiencia de la “inarticulación” y, con ello, a la imposibilidad de dar y escuchar razones, si se insistía en su idea de que las decisiones que se tomen ante los dilemas morales obedecerían a una necesidad “volitiva” en la cual Frankfurt *no* cree que estén involucrados elementos cognitivos. Si bien habría en este autor una tesis con la cual Taylor estaría de acuerdo, a saber, la tesis de que aquello que constituye las preocupaciones fundamentales de una persona no es objeto de la mera elección por parte de ésta, sin embargo, también se vio que la insistencia de Frankfurt en que no cabría apelar a referentes normativos para la justificación de las decisiones que el agente tome con miras a no traicionar dichas preocupaciones, le sumirían a éste en una suerte de mutismo que cerraría las puertas a lo cognitivo. Con ello, el agente de Frankfurt, al igual que el de Gauthier, tendría que ser pensado bajo un modelo monológico que, como se vio en los primeros tres capítulos, implicaría concebirlo como el Robinson solitario, sin vínculos con otros agentes y sin un acervo de significados compartidos con esos otros. Con lo cual, en su caso, y de modo similar a lo que le ocurre al *homo oeconomicus* de Gauthier, también se “cerraría” su base motivacional ante la crítica, ante el dar y escuchar razones. Sin embargo, la persona de Frankfurt tendría el plus de la profundidad, de la reflexión, de la autoevaluación y del esfuerzo que esta última conlleva.

Con esto, pues, volvemos al punto de inicio del presente trabajo: ¿qué motivos habría para estar preocupados por las consecuencias que traería el asumir como modélica una forma monológica de agencia? Tratándose del *homo oeconomicus*, creo que la preocupación a la que me refiero apunta a las limitaciones que este último impondría si se lo asume como modelo de la agencia humana, limitaciones que tal vez podrían ser resumidas en una incapacidad para la autocrítica y para la reflexión. Creo que esto sería un motivo válido de preocupación por razones que acaso cabría abreviar

bajo la expresión “pérdida de libertad”. Pues quien se pensara a sí mismo y pensara a sus congéneres como siendo básicamente unos maximizadores egoístas, se autoimpondría e impondría a los demás ciertos límites para la reflexión y la discusión, para el despliegue de las capacidades deliberativas y evaluativas, para el dar y pedir razones, límites que conllevarían la instalación de un mundo social cuyos mecanismos no cabría criticar ni intervenir. Estos límites, en el caso de cada agente, significarían que su acción obedecería a una base motivacional estrecha y opaca, el autointerés, el cual, a su vez, al estar atado al uso excluyente de una forma maximizadora-instrumental de racionalidad práctica, también parecería estar blindado a todo intento de aclaración, de desvelamiento y de examen que permitiese “abrirlo” y exponerlo en su complejidad; complejidad que no cabría incluir en una fórmula tan escueta como la que designan expresiones tales como “autointerés”, “beneficio”, “utilidad”.

Dicha complejidad de los deseos, propósitos, fines e ideales humanos, de ser desplegada y ganada para la reflexión, el autoexamen, y la crítica que puedan llevar a cabo, tanto el propio sujeto, como aquellos agentes con los cuales converse sobre las claves de significación y de normatividad de tal base motivacional, también implicaría más posibilidades de reconstruir el sentido de los actos y de las interacciones humanas y, con ello, una mayor autoconciencia, y una mayor apertura a las diversas posibilidades para la acción y la autointerpretación, las cuales no serían alcanzables por parte de un sujeto al que se conciba como insensible a las voces de sus congéneres. Esta apertura y esta conciencia irían, pues, de la mano de la escucha de lo que otros tuvieran qué decir sobre las motivaciones que cada agente señale como la base de su acción, así como sobre la deseabilidad y/o legitimidad de las relaciones que se establezcan entre las personas, para lo cual se haría necesaria la apelación a referentes de sentido y de normatividad que, de suyo, no podrían ser articulados en términos que sean inteligibles y aceptables únicamente para el agente en solitario, sino que se corresponderían con los sentidos compartidos o, al menos, compartibles con otros.

Esto último, a su vez, señala la posibilidad de unos vínculos intersubjetivos mucho más ricos y abiertos a la discusión, en contraste con aquellos otros que puedan desarrollar entre sí unos agentes a los que se piense, o que se piensen a sí mismos y a sus congéneres, en clave meramente estratégica; es decir, como obstáculos o como medios para unos fines incuestionados, o como seres con los que se coopera únicamente

por razones ancladas en esa base motivacional opaca y estrecha que se resume en la escueta fórmula del “autointerés”. Por el contrario, si los vínculos intersubjetivos estuviesen abiertos a la crítica, a la revisión, también se contaría con la posibilidad de una mayor libertad en el sentido que comportarían unas relaciones interpersonales que no se impongan mediante el uso de la fuerza o de la manipulación o que, al menos, puedan ser cualificadas por la vía de su aclaración, discusión y crítica.

Esto último se vio como una posibilidad más clara en el modelo del evaluador fuerte de Taylor, si bien, así mismo se cayó en cuenta de que dicha libertad tendría que ser pensada como un fin, como un logro deseable y necesario, pero no como algo que el agente tenga asegurado desde el principio. ¿Cómo concebir, entonces, las condiciones de dicho logro? La respuesta a esta pregunta indicaría los posibles caminos a seguir por la investigación que se inició con el presente trabajo. Uno de esos caminos que, por lo pronto, se alcanza a vislumbrar, apuntaría al asunto de la cualificación de eso que Taylor llama “autointerpretaciones”, y al papel que en ellas puedan desempeñar las voces de los otros, de esos agentes de quienes aprendimos el lenguaje que usamos para hablar de nosotros mismos. En este sentido, la escena que muestra Taylor de un agente que lucha por autoaclararse, por aclarar sus compromisos valorativos más importantes, creo que necesitaría complementarse con una descripción más detallada de los modos en los que se hacen presentes, en ese diálogo del agente consigo mismo, las voces de esos otros con los cuales la persona comparte los referentes de valor a los que apela en sus intentos por articular lo que las cosas significan para ella.

Uno de dichos modos tal vez se relacionaría con aquello que proporciona parte de las “razones” que podría ofrecer el agente para su acción: su propia historia, lo que lo ha llevado hasta donde está. Si la persona no podría narrar esa historia sin incorporar en ella a los otros, a esos otros cuyas historias se entrelazan con la del propio agente, creo que habría que intentar, no sólo una narración-articulación de dicha historia, sino también una *crítica* de ésta, así como de los referentes de sentido y de significación a los que se apele para llevar a cabo tal narración. Ello, con el fin de que dichos referentes y dicha historia no se conviertan en “hechos” innegables, incuestionables y, por tanto, para que no se merme la libertad de la persona, en el sentido de cerrar para ella las opciones de autointerpretación y, por ende, las opciones de elección y de acción. Pues es posible que aquello a lo cual el agente se (auto)descubra como fundamentalmente

atado o comprometido, también necesite de un examen crítico, que si bien requeriría de una conversación con otros, nada nos asegura que dicha conversación sea transparente, libre de manipulaciones y libre de diversas formas de distorsión.

De este modo, es posible que la lucha del agente por ganarse a sí mismo, mediante sus intentos de autoaclaración, cuente con un elemento más de complejidad y de dificultad, que vendría dado por formas desafortunadas en las que las voces de otros se hayan hecho presentes distorsionando, bien la autoimagen del propio agente, bien la idea que éste pueda hacerse de su mundo social. Pensemos, además, en que es posible que esta conversación *no* tenga lugar en un escenario caracterizado por la simetría y, por ende, en donde no todos los que allí conversan tengan las mismas posibilidades de hacer valer su voz y su mirada, tanto para elaborar el relato que cada quien pueda ofrecer sobre sí mismo, como para influenciar el relato de los demás, o influenciar “la” historia que se asuma como la de todos o que contemos como la de “todos”. Por lo anterior, pienso que una de las rutas a seguir apuntaría hacia una exploración de los modos en los que podría advertirse este “peso” que tendrían sobre la propia autoimagen de las personas, el reconocimiento o la falta de reconocimiento, por parte de sus congéneres, habida cuenta de que los escenarios sociales reales, en los cuales transcurre la vida de esas personas, no necesariamente se caracterizan por la simetría. De tal modo que su autoimagen estaría necesitada de una evaluación, de un proceso de crítica que permitiría el hacer concientes, el abrir a la discusión los elementos simbólicos que inciden en la construcción de dicha autoimagen. Estos elementos, a su vez, incidirían en las posibilidades para una libertad entendida en clave de “oportunidad”, de apertura de opciones para pensarnos y, con ello, para que se nos muestren las posibilidades, las diversas opciones para la autovaloración, para la valoración de otras formas de vida, así como para la acción y para las posibles formas en que podríamos construir nuestros lazos con las otras personas.

Bibliografía

- Abell, Peter (ed.). (1991). *Rational Choice Theory*. Cambridge: Cambridge University Press.
- Apel, Karl Otto. (1985). *La transformación de la filosofía*. Madrid: Taurus.
- _____. (1991). *Teoría de la verdad y ética del discurso*. Barcelona: Paidós.
- Appiah, K. Anthony. (2007). *La ética de la identidad*. Buenos Aires: Katz.
- Aristóteles. *Ética a Nicómaco*. Madrid: Centro de Estudios Constitucionales, 1959. (Trad. Julián Marías y María Zambrano).
- Arrow, Kenneth J. (1996). "Social Responsibility and Economic Efficiency", en: Donaldson, Thomas y Werhane, Patricia (Eds.). *Ethical Issues in Business: A Philosophical Approach*, pp. 227-237.
- Axelrod, Robert. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- _____. (1997). *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton (N. J.): Princeton University Press.
- Baron, Jonathan. (1993). *Morality and Rational Choice*. Dordrecht: Kluwer Ac. Pub.
- Bator, Francis M. (1958). "The Anatomy of Market Failure", en: *The Quarterly of Journal of Economics*, pp. 351-379.
- Brandom, Robert. (1998). "Actions, Norms, and Practical Reasoning", en: *Philosophical Perspectives*, 12, (Supplement: *Language, Mind, and Ontology*), pp. 127-139.
- _____. (2001). "What do Expressions of Preferences Express?", en: Morris, Christopher and Ripstein, Arthur (eds.), *Practical Rationality and Preference. Essays for David Gauthier*, pp. 11-36.
- Bratman, Michael E. (2002). "Hierarchy, Circularity, and Double Reduction", en: Buss, Sarah, y Overton, Lee (eds). (2002). *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 65-85.
- _____. (2006). "A Thoughtful and Reasonable Stability", en: Frankfurt, Harry, *Taking Ourselves Seriously, and Getting it Right*, pp. 77-90.
- Breit, William and Hochman, Harold M. (1971). *Readings in Microeconomics*. Hinsdale (Illinois): Dryden Press.
- Broncano, Fernando. (1988). "Racionalidad e ideología. ¿Cómo ser un buen realista con respecto de las ideologías?", en: Olivé, León (ed.), *La racionalidad. Ensayos sobre la racionalidad en ética y política, ciencia y tecnología*, pp. 176-194.
- Buss, Sarah, y Overton, Lee (eds). (2002). *Contours of Agency. Essays on Themes from Harry Frankfurt*. Cambridge (MA): MIT Press.

- Butler, Judith. (2005). *Giving an Account of Oneself*. New York: Fordham University Press.
- Catanzaro, Raimondo. (1985). "Enforcers, Entrepreneurs, and Survivors: How the Mafia Has Adapted to Change", en: *The British Journal of Sociology*, Vol. 36, No. 1, pp. 34-57.
- Cerezo, Pedro. (ed.). *Democracia y virtudes cívicas*. Madrid: Biblioteca Nueva.
- Davidson, Donald. (2001). "Actions, Reasons, and Causes", en: *Essays on Actions and Events*, pp. 3-20.
- _____. (2001a). *Essays on Actions and Events*. New York: Oxford University Press.
- De Sousa, Ronald. (1980). "Self-deceptive Emotions", en: Rorty, Amelie Oksenberg (Ed.), *Explaining Emotions*, pp. 283-298.
- Donaldson, Thomas y Werhane, Patricia (eds.). (1996). *Ethical Issues in Business Ethics: A Philosophical Approach*. New Jersey: Prentice-Hall.
- Elster, Jon (1995). *Ulises y las sirenas: estudios sobre racionalidad e irracionalidad*. México: Fondo de cultura económica.
- _____. (2002). *Ulises desatado*. Barcelona: Gedisa.
- Flanagan, Owen y Rorty, Amélie Oksenberg (eds.). (1993). *Identity, Character and Morality: Essays in Moral Psychology*. Cambridge: MIT Press.
- Fodor, Jerry. (1985). *El lenguaje del pensamiento*. Madrid: Alianza. (Trad. Jesús Fernández Zulaica).
- Frankfurt, Harry G. (1998). "Freedom of the will and the concept of a person", en: *The Importance of what we Care about. Philosophical Essays*, pp. 11-25.
- _____. (1998a). "Identification and externality", en: *The Importance of what we Care about. Philosophical Essays*, pp. 58-68.
- _____. (1998b). "Identification and wholeheartedness", en: *The importance of what we Care About*, pp. 159-176.
- _____. (1998c). "Rationality and the unthinkable", en: *The importance of what we Care about*, pp. 177-190.
- _____. (1998d). "The importance of what we Care About", en: *The Importance of what we Care about. Philosophical Essays*, pp. 80-94.
- _____. (1998e). "Three concepts of free action" ", en: *The Importance of what we Care about. Philosophical Essays*, pp. 47-57.
- _____. (1998f). *The Importance of what we Care about. Philosophical Essays*. New York: Cambridge University Press.
- _____. (2002). "Replay to Gary Watson", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 160-164.

- _____. (2002a). "Replay to J. David Velleman", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 124-128.
- _____. (2002b). "Replay to Jonathan Lear", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 293-299.
- _____. (2002c). "Replay to Michael Bratnam", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 86-90.
- _____. (2002d). "Replay to Richard Moran", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 218-226.
- _____. (2002e). "Replay to Susan Wolf", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 245-252.
- _____. (2002f). "Replay to T. M. Scanlon", en: Buss, Sarah, y Overton, Lee (eds), *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 184-188.
- _____. (2004). *Las razones del amor. El sentido de nuestras vidas*. Barcelona: Paidós.
- _____. (2006). *Taking Ourselves Seriously, and Getting it Right*. Stanford (CA): Stanford University Press.
- _____. (2007). "Acerca de la libertad y los límites de la voluntad", en: *Necesidad, volición y amor*, pp. 119-133. (Título original: "On Freedom and the Limits of the Will").
- _____. (2007a). "Autonomía, necesidad y amor", en: *Necesidad, volición y amor*, pp. 205-222. (Título original: "Necessity, Volition and Love").
- _____. (2007b). "La más tenue de las pasiones", en: *Necesidad, volición y amor*, pp. 155-173. (Título original: "The Faintest Passion").
- _____. (2007c). "Sobre el preocuparse", en: *Necesidad, volición y amor*, pp. 243-278. (Título original: "On Caring").
- _____. (2007d). *Necesidad, volición y amor*. Buenos Aires: Katz.
- Freeman, Samuel R. (2007). *Justice and the Social Contract. Essays on Rawlsian Political Philosophy*. Oxford: Oxford University Press.
- _____. (2007a). "Reason and Agreement in Social Contract Views", en: *Justice and the Social Contract. Essays on Rawlsian Political Philosophy*, pp. 17-44.
- _____. (2007b). *Rawls*. New York: Routledge.
- _____. (Ed.) (2003). *The Cambridge Companion to Rawls*. New York: Cambridge University Press.
- Friedman, Milton. (1966). *Capitalismo y Libertad*. Madrid: Rialp.
- _____. (1980). *Libertad de elegir. Hacia un nuevo liberalismo económico*. Barcelona: Grijalbo.

- _____. (1996). "The Social Responsibility of Business Is to Increase Its Profits", en: Donaldson, Thomas y Werhane, Patricia (Eds.). *Ethical Issues in Business: A Philosophical Approach*, pp. 222-227.
- Fukuyama, Francis. (1992). *El fin de la historia y el último hombre*. Barcelona: Planeta.
- _____. (1996). *Trust. The Social Virtues and the Creation of Prosperity*. London: Penguin Books.
- Gambetta, Diego. (1993). *The Sicilian Mafia: The Business of Private Protection*. Cambridge (Ma): Harvard University Press.
- Gauthier, David y Sugden, Robert (eds.). (1993). *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*. New York: Harvester Wheatsheaf.
- Gauthier, David. (1963). *Practical Reasoning. The Structure and Foundations of Prudential and Moral Arguments and, their Exemplifications in Discourse*. Oxford: Clarendon Press.
- _____. (1986). *Morals by Agreement*. Oxford: Clarendon Press. (Versión en español: *La moral por acuerdo*. Barcelona: Gedisa, 2000. Trad. Alcira Bixio).
- _____. (1993). "Between Hobbes and Rawls". En: Gauthier, David y Sugden, Robert (eds.). *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*, pp. 24-39.
- _____. (1998) *Egoísmo, moralidad y sociedad liberal*. Barcelona: Paidós.
- _____. (1998a). "El egoísta incompleto", en: Gauthier, David, *Egoísmo, moralidad y sociedad liberal*, pp. 67-117.
- Geuss, Raymond. (1996). "Morality and Identity", en: Korsgaard, Christine M, *The Sources of Normativity*, pp. 189-199.
- Gibson, Mary. (1977). "Rationality", en: *Philosophy and Public Affairs*, Vol. 6, No. 3, pp. 193-225.
- Goodin, Robert E. (1993). "Equal Rationality and Initial Endowments", en: Gauthier, David y Sugden, Robert (eds.), *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*, pp. 116-130.
- Graebner Anderson, Annelise. (1979). *The Business of Organized Crime. A Cosa Nostra Family*. Stanford: Hoover Institution Press.
- Greenspan, Patricia S. (1980). "A Case of Mixed Feelings: Ambivalence and the Logic of Emotion", en: Rorty, Amelie Oksenberg (Ed.), *Explaining Emotions*. Berkeley: University of California Press, pp. 223-250.
- Habermas, Jürgen. (1987). *Teoría de la acción comunicativa*. Madrid: Taurus.
- _____. (1989). "Aspectos de la racionalidad de la acción", en: *Teoría de la acción comunicativa: complementos y estudios previos*, pp. 369-395.

- _____. (1989a). *Teoría de la acción comunicativa: complementos y estudios previos*. Madrid: Cátedra.
- _____. (1990). "Acerca del uso ético, pragmático y moral de la razón práctica", en: *Filosofía*, No. 1, Mérida (Venezuela), pp. 7-24.
- _____. (1994). "Conciencia moral y acción comunicativa", en: *Conciencia moral y acción comunicativa*. Barcelona: Península, 1991, pp. 135-219.
- _____. (1994a). "Ética del discurso. Notas sobre un programa de fundamentación", en: *Conciencia moral y acción comunicativa*, pp. 98-134.
- _____. (1994b). *Conciencia moral y acción comunicativa*. Barcelona: Península,
- Hahn, Frank, y Hollis, Martin. (1992). (Eds.). *Filosofía y teoría económica*. México: Fondo de cultura económica.
- Hampton, Jean (1988). "Forgiveness, Hatred and Mercy", en: Hampton, Jean y Murphy, Jeffrie, *Forgiveness and Mercy*, pp. 35-87.
- Hampton, Jean y Murphy, Jeffrie. (1988). *Forgiveness and Mercy*. Cambridge: Cambridge University Press.
- Harsanyi, John C. (1982). "Morality and the Theory of Rational Behavior", en: Sen, Amartya y Williams, Bernhard (eds.), *Utilitarianism and Beyond*, pp. 39-62.
- _____. (1982a). *Papers in Game Theory*. Dordrecht: D. Reidel Publishing.
- _____. (1983). "Basic Moral Decisions and Alternative Concept of Rationality", en: *Social Theory and Practice*, 9: 2/3, pp. 231-244.
- _____. (1985). "Does Reason Tell Us What Moral Code to Follow and, indeed, to Follow Any Moral Code at All?", en: *Ethics*, 96, pp. 42-55.
- _____. (1988). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
- Hayek, Friedrich A. Von. (1978). *Camino de servidumbre*. Madrid: Alianza.
- _____. (1991). *Economic Freedom*. Oxford: Blackwell.
- _____. (1998). *Los fundamentos de la libertad*. Madrid: Unión editorial.
- Heller, Agnes. (1990). *Más allá de la Justicia*. Barcelona: Crítica.
- Hobbes, Thomas. *Leviatán. La material, la forma y el poder en un estado eclesiástico y civil*. Madrid: Alianza, 1993. (Trad. Carlos Mellizo).
- Hume, David. *Investigación sobre los principios de la moral*. Madrid: Alianza, 1993. (Trad. Carlos Mellizo).
- _____. *Tratado de la naturaleza humana*. Madrid: Tecnos, 1988. (Trad. Félix Duque).

- Huxley, Aldous. (1989). *Un mundo feliz*. Barcelona: Plaza & Janés. (Edición original: 1969).
- Kant, Immanuel. *Ideas para una historia universal en clave cosmopolita y otros escritos sobre Filosofía de la Historia*. Madrid: Tecnos, 2006. (Trad. Concha Roldán Panadero y Roberto Rodríguez Aramayo).
- Korsgaard, Christine M. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press. Versión en español: *Las fuentes de la normatividad*. México: Universidad Nacional Autónoma de México, Instituto de Investigaciones Filosóficas, 2000.
- _____. (2006). "Morality and the Logic of Caring", en: Frankfurt, Harry, *Taking Ourselves Seriously, and Getting it Right*, pp. 55-76.
- _____. (2009). *Self-Constitution. Agency, Identity, and Integrity*. New York: Oxford University Press.
- Larmore, Charles. (2003). "Public Reason", en: Freeman, Samuel R. (Ed.) *The Cambridge Companion to Rawls*, pp. 368-393.
- Locke, John. *Segundo tratado sobre el gobierno civil. Un ensayo acerca del verdadero origen, alcance y fin del gobierno civil*. Madrid: Alianza, 1996. (Trad. Carlos Mellizo).
- Luce, Duncan R. y Raiffa, Howard. (1958). *Games and Decisions. Introduction and Critical Survey*. New York: John Wiley and Sons.
- MacIntyre, Alasdair. (1987). *Tras la virtud*. Barcelona: Crítica.
- _____. (1991). *Historia de la ética*. Barcelona: Paidós.
- Mackie, J. L. (1990). *Ethics: Inventing Right and Wrong*. New York: Penguin Books. (Versión española: *Ética: la invención de lo bueno y lo malo*, Barcelona, Gedisa, 2000, Trad. Tomás Fernández Auz).
- Macpherson, C. B. (1962). *The Political Theory of Possessive Individualism. Hobbes to Locke*. Oxford: Oxford University Press.
- Máiz, Ramón (comp). (2009). *Teorías políticas contemporáneas*. Valencia: Tirant lo blanch.
- Mathiot, Jean. (1990). *Adam Smith: philosophie et économie: de la sympathie à l'échange*. Paris: Presses Universitaires de France.
- Mead, George Herbert. (1972). *Mind, Self, and Society. From the Standpoint of a Social Behaviorist*. Chicago: Chicago University Press. (First published 1934). Versión en español: (1990). *Espíritu, persona y sociedad: desde el punto de vista del conductismo social*. Barcelona: Paidós.
- Moore, G. E. *Ensayos éticos*. Barcelona: Paidós, 1993.
- _____. *Ética y análisis*. México: UNAM, 1985.
- _____. *Ética*. Barcelona: Labor, 1989. (Trad. Manuel Cárdenas Iracheta).
- _____. *Principia Etica*. México: UNAM, 1997. (Trad. Thomas Baldwin).

- Moran, Richard. (2002). "Frankfurt on Identification: Ambiguities on Activities in Mental Life", en: Buss, Sarah, y Overton, Lee (eds). (2002). *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 189-217.
- Morris, Christopher and Ripstein, Arthur (eds.). (2001). *Practical Rationality and Preference. Essays for David Gauthier*. Cambridge: Cambridge University Press.
- Murphy, Jeffrie. (1988). "Forgiveness and Resentment", en: Hampton, Jean y Murphy, Jeffrie, *Forgiveness and Mercy*, pp. 14-34.
- North, Douglas. (1994). *Instituciones, cambio institucional y desarrollo económico*. México: Fondo de cultura económica.
- Nozick, Robert. (1990). *Anarquía, estado y utopía*. México: Fondo de cultura económica.
- _____. (1995). *La naturaleza de la racionalidad*. Barcelona: Paidós.
- Olivé, León (ed.). (1988). *La racionalidad. Ensayos sobre la racionalidad en ética y política, ciencia y tecnología*. México: Siglo XXI.
- Ovejero Lucas, Félix. (1994a). "Las defensas morales del mercado", en: *Isegoría*, No. 9, pp. 41-63.
- _____. (1994b). *Mercado, ética y economía*. Barcelona: Icaria-Fuhen.
- _____. (1998). "Del mercado al instinto (o de los intereses a las pasiones)", en: *Isegoría*, No. 18, pp. 181-204.
- _____. (2002). *La libertad inhóspita: modelos humanos y democracia liberal*. Barcelona: Paidós.
- _____. (2004). "Economía y psicología: entre el método y la teoría", en: *Revista internacional de sociología*, No. 38, p. 9-34.
- Parfit, Derek. (1984). *Reasons and Persons*. Oxford: Oxford University Press. Versión en español: 2004, *Personas, racionalidad y tiempo*. Madrid: Síntesis.
- Pettit, Phillip. (1986). "Free Riding and Foul Dealing", en: *The Journal of Philosophy*. Vol. 83, No. 7. pp. 361-379.
- Platón. *La República*. Madrid: Centro de Estudios Constitucionales, 1981. (Trad. José Manuel Pabón y Manuel Fernández Galiano).
- Pogge, Thomas W. M. (2007). *John Rawls: His Life and the Theory of Justice*. New York. Oxford University Press.
- Putnam, Hilary. (1998). "Racionalidad en la teoría de la decisión y en la ética", en: Olivé, León (ed.), *Racionalidad: ensayos sobre racionalidad en ética y política, ciencia y tecnología*, pp. 46-57.
- _____. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge: Harvard University Press.

- Rapaport, Anatol y Chammah, Albert. (1965). *Prisoner's Dilemma. A Study in Conflict and Cooperation*. Ann Arbor: The University of Michigan Press.
- Rawls, John. (1986). *Justicia como equidad. Materiales para una teoría de la justicia*. Madrid: Tecnos.
- _____. (1996). *Liberalismo político*. México, Fondo de cultura económica.
- _____. (1999). *A Theory of Justice*. Cambridge (MA): Harvard University Press. (First edition 1971). Versión en español: *Teoría de la justicia*. México, Fondo de cultura económica, 1995.
- Rescher, Nicholas. (1993). *La racionalidad. Una indagación filosófica sobre la naturaleza y la justificación de la razón*. Madrid: Tecnos.
- Resnik, Michael D. (1998). *Elecciones. Una introducción a la teoría de la decisión*. Barcelona: Gedisa.
- Ripstein, Arthur. (2001). "Preference", en: Morris, Christopher and Ripstein, Arthur (eds.), *Practical Rationality and Preference. Essays for David Gauthier*, pp. 37-55.
- Rorty, Amelie Oksenberg (Ed.). (1980). *Explaining Emotions*. Berkeley: University of California Press.
- _____. (1980a). "Explaining Emotions", en: *Explaining Emotions*, pp. 103-126.
- Rousseau, Jean-Jacques. (2001). *Discurso sobre el origen y los fundamentos de la desigualdad entre los hombres; y otros escritos*. Madrid: Tecnos.
- Ryle, Gilbert. (2002). *The Concept of Mind*. Chicago: University of Chicago Press. En español: *El concepto de lo mental*. Barcelona: Paidós, 2005. (Trad. Eduardo Rabossi).
- Sartre, Jean Paul. (1978). *El existencialismo es un humanismo*. Buenos Aires: Sur.
- _____. (1979). *El ser y la nada*. Buenos Aires: Losada.
- Scanlon, Thomas M. (2002). "Reasons and Passions", en: Buss, Sarah, y Overton, Lee (eds). (2002). *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 164-183.
- _____. (2003). "Rawls on Justification", en: Freeman, Samuel R. (Ed.) *The Cambridge Companion to Rawls*, pp. 139-167.
- Scheler, Max. (1941). *Ética: nuevo ensayo de fundamentación de un personalismo ético*. Madrid: Revista de Occidente. (Trad. Hilario Rodríguez Sanz).
- _____. (2007). *De lo eterno en el hombre*. Madrid: Encuentro. (Trad. Julián Marías y Javier Olmo).
- Schmidtz, David. (2006). *Rational Choice and Moral Agency*. Princeton University Press.
- Sen, Amartya y Williams, Bernhard (eds.). (1982). *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

- Sen, Amartya. (1992). "Los tontos racionales: una crítica de los fundamentos conductistas de la teoría económica", en: Hahn, Frank, y Hollis, Martin (eds.), *Filosofía y teoría económica*, pp. 173-217. (Título original: "Rational Fools: A Critique of the Behavioural Foundations of Economic Theory", en: *Philosophy and Public Affairs*, Vol. 6, No. 4, 1977, pp. 317-344).
- _____. (1997). "Maximization and the Act of Choice". En: *Econometrica*, Vol. 65, No. 4, pp. 745-779.
- _____. (1999). *Development as Freedom*. Oxford: Oxford University Press.
- _____. (2003). *Rationality and Freedom*. Cambridge: Harvard University Press.
- Smart, J. J. C. (1988). "An Outline of a System of Utilitarian Ethics", en: Smart, J.J.C y Williams, Bernard, *Utilitarianism For and Against*, pp. 3-74.
- Smart, J. J. C. y Williams, Bernard. (1988). *Utilitarianism For and Against*. Cambridge: Cambridge University Press.
- Smith, Adam. *Investigación sobre la naturaleza y las causas de la riqueza de las naciones*. México: Fondo de cultura económica, 1982. (Trad. Gabriel Franco).
- _____. *Teoría de los sentimientos morales*. Madrid: Alianza, 2004. (Trad. Carlos Rodríguez Braun).
- Smith, Dwight C. Jr. (1975). *The Mafia Mystique*. New York: Basic Books.
- Sobrevilla, David (comp). (1991). *El derecho, la política y la ética*. México: Siglo XXI.
- Stevenson, Charles L. (1984). *Ética y lenguaje*. Barcelona: Paidós. (Trad. Eduardo Rabossi).
- Stocker, Michael. (1980). "Intellectual Desire, Emotion, and Action", en: Rorty, Amelie Oksenberg (Ed.), *Explaining Emotions*, pp. 223-238.
- Strawson, Peter F. (1995). "Libertad y resentimiento", en. *Libertad y resentimiento, y otros ensayos*. Barcelona: Paidós, pp. 37-67.
- _____. (1995a). *Libertad y resentimiento, y otros ensayos*. Barcelona: Paidós.
- Stroud, Barry. (1986). *Hume*. México: UNAM.
- Suits, Bernard. (1978). *The Grasshopper: Games, Life and Utopia*. Toronto: University of Toronto Press.
- Taylor, Charles. (1985). "Rationality", en: *Philosophy and the Human Sciences. Philosophical Papers II*, pp. 134-151.
- _____. (1985a). "Self-Interpreting Animals", en: *Human Agency and Language. Philosophical Papers I*, pp. 45-76.
- _____. (1985b). "The Concept of a Person", en: *Human Agency and Language. Philosophical Papers I*, pp. 97-114.

- _____. (1985c). "What is Human Agency?", en: *Human Agency and Language. Philosophical Papers I*, pp. 15-44.
- _____. (1985d). *Human Agency and Language. Philosophical Papers I*. Cambridge: Cambridge University Press.
- _____. (1985e). *Philosophy and the Human Sciences. Philosophical Papers II*. Cambridge: Cambridge University Press.
- _____. (1994). *La ética de la autenticidad*. Barcelona: Paidós.
- _____. (1996). *Fuentes del yo*. Barcelona: Paidós.
- _____. (1997). *Argumentos Filosóficos. Ensayos sobre el conocimiento, el lenguaje y la modernidad*. Barcelona: Paidós.
- _____. (1997a). "La explicación y la razón práctica", en: *Argumentos Filosóficos. Ensayos sobre el conocimiento, el lenguaje y la modernidad*, pp. 59-90.
- _____. (2006). *Imaginarios sociales modernos*. Barcelona: Paidós.
- _____. (2007). *A Secular Age*. Cambridge (MA): Harvard University Press.
- Thiebaut, Carlos. (1994). "Recuperar la moral: la filosofía de Charles Taylor", estudio introductorio a: Taylor (1994), pp. 11-34.
- _____. (2005). "La prudencia", en: Cerezo, Pedro (ed.), *Democracia y virtudes cívicas*, pp. 75-111.
- _____. (2008). "El arrepentimiento: un caso pro y contra de la (dis)continuidad del yo". Manuscrito.
- _____. (2009). "Charles Taylor: democracia y reconocimiento", en: Máiz, Ramón (Comp.) *Teorías políticas contemporáneas*, pp. 211-232.
- Tugendhat, Ernst. (1988). "El acceso semántico a la moral", en: *Problemas de la ética*, pp. 66-99.
- _____. (1988a). *Problemas de la ética*. Barcelona: Crítica.
- _____. (1991) "Una nueva concepción de la filosofía moral", en: Sobrevilla, David (comp), *El derecho, la política y la ética*, pp. 151-173.
- Velleman, J. David. (2002). "Identification and Identity", en: Buss, Sarah, y Overton, Lee (eds). *Contours of Agency. Essays on Themes from Harry Frankfurt*, pp. 91-123.
- Von Neumann, John y Morgenstern, Oskar. (1953). *Theory of Games and Economic Behaviour*. Princeton: Princeton University Press.
- Watson, Gary. (1975). "Free Agency", en: *Journal of Philosophy*, vol. 72, No. 8, pp. 205-220.

- _____. (2002). "Volitional Necessities", en: Buss, Sarah, y Overton, Lee (eds). Williams, Bernard. (1995). "Internal Reasons and the Obscurity of Blame", en: *Making Sense of Humanity, and other Philosophical Papers, 1982-1993*, pp. 35-45.
- Wellmer, Albrecht. (1988). "Intersubjetividad y razón", en: Olivé, León (ed.), *La racionalidad. Ensayos sobre la racionalidad en ética y política, ciencia y tecnología*, pp. 225-266.
- Williams, Bernard. (1972). *Morality*. New York: Cambridge University Press.
- _____. (1988). "A Critique of Utilitarianism", en: Smart, J.J.C y Williams, Bernard, *Utilitarianism for and against*, pp. 77-150.
- _____. (1995). "Internal Reasons and the Obscurity of Blame", en: *Making Sense of Humanity, and other Philosophical Papers, 1982-1993*, pp. 35-45.
- _____. (1995a). *Making Sense of Humanity, and other Philosophical Papers, 1982-1993*. Cambridge: Cambridge University Press.
- _____. (2002). *Truth and Truthfulness. An Essay in Genealogy*. Princeton (NJ): Princeton University Press.
- Wittgenstein, Ludwig. (2002). *Investigaciones filosóficas*. Barcelona: Crítica. (Trad. Alonso García Suárez y Ulises Moulines. Edición bilingüe; segunda edición).
- Young, Iris Marion. (1990). *Justice and the Politics of Difference*. Princeton: Princeton University Press.
- Young, Jeffrey T. (1997). *Economics as a Moral Science: the Political Economy of Adam Smith*. Cheltenham: Edward Elgar.